SAXS AND X-RAY CRYSTALLOGRAPHIC STUDIES OF THE CELLULOSOME FROM CLOSTRIDIUM THERMOCCELLUM

by

Mark Allister Currie

A thesis submitted to the Department of Biochemistry
In conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
(August, 2010)

Copyright ©Mark Allister Currie, 2010
Abstract

Cellulosomes are the most efficient plant cell wall degradation machines discovered to date. All cellulosomal components contain protein modules connected by linkers of varying lengths, which are predicted to be flexible. Consequently, structural studies of the cellulosome have employed a “dissect and build” strategy, whereby individual modules are studied in isolation with the hope to later model the intact complex. However, representative individual structures are now available for all of the cellulosome modules and many questions still remain. The studies described in this thesis depart from the ‘dissection’ stage and enter the ‘build’ stage of cellulosome structural studies of the cellulosome from *Clostridium thermocellum*.

We first describe the crystal structure of a heterodimeric complex comprising the type-II cohesin (CohII) from cell surface anchoring protein SdbA and a trimodular C-terminal truncation of the CipA scaffoldin protein containing the ninth type-I cohesin module (CohI$_9$), a linker, the X-module (X), and the type-II dockerin module (DocII). This structure revealed novel intertwining of scaffoldin molecules and extensive reciprocal contacts between the CohI$_9$ and the X-module of another scaffoldin molecule. Sedimentation velocity experiments indicate dimerization also occurs in solution.

We have carried out the crystallization and structural analysis of a heterotrimeric complex containing the CohI$_9$—X-DocII:CohII complex bound to the type-I dockerin module (DocI) from the Cel9D enzyme, which represents the largest cellulosome fragment ever determined. Identical inter-scaffoldin interactions were observed in the heterotrimeric complex structure as were seen in the heterodimeric complex. However,
small angle X-ray scattering (SAXS) studies indicate that this dimerization does not occur in solution. The crystal structures and additional SAXS studies reveal flexibility in the CohI$_9$—X linker that is surprisingly restricted to two dimensions. In addition, this structure provides the first evidence of an orientation bias in DocI binding.

Finally, SAXS was used to investigate modular orientations and linker flexibility in several cellulosome fragments. These studies indicate that cellulosomal linkers exhibit restricted and in some cases highly restrained flexibility. Specifically, scaffoldin linkers display two dimensional motions, enzymes maintain close contact with their cognate DocI modules, and enzyme positions rotate about 90° relative to neighbouring enzymes on the scaffoldin.
Co-Authorship

This section describes the contributions made by each author to the experimental work and manuscript preparation presented in each chapter.

**Chapter 2:** This paper was published in the Journal of Molecular Biology in March 2010. The authors are Jarrett J. Adams, Mark A. Currie, Edward A. Bayer, Zongchao Jia, and Steven P. Smith. Jarrett Adams solved the structure presented and performed the analytical ultracentrifugation experiments. I performed DLS and SAXS experiments that were ultimately omitted from the final manuscript. I also performed structural analysis, wrote the initial draft of the manuscript, and participated in editing the manuscript along with Steven Smith, Edward Bayer, and Zongchao Jia.

**Chapter 3:** This paper was published in Acta Crystallographica, Section F, Structural Biology and Crystallization Communications in March 2010. The authors are Mark A. Currie, Jarrett J. Adams, Steven P. Smith, and Zongchao Jia. Jarrett Adams cloned the Coh19—X-DocII fragment from a *Clostridium thermocellum* CipA construct obtained from Dr. Pierre Bèguin (Pasteur Institute, Paris) and identified the initial crystallization condition. I subsequently optimized the crystallization conditions, cryoprotectant, and collected and analyzed X-ray diffraction data. I wrote the initial draft of the manuscript and participated in subsequent editing with Steven Smith and Zongchao Jia.

**Chapter 4:** This manuscript is unpublished work carried out during my Ph.D. The authors are Mark A. Currie, Jarrett J. Adams, Edward A. Bayer, Steven P. Smith, and Zongchao Jia. I performed all of the experiments described in this manuscript. Jarrett
Adams assisted with structure analysis. I wrote the initial draft of the manuscript and participated in subsequent editing with Jarret Adams, Steven Smith, Edward Bayer, and Zongchao Jia.

**Chapter 5:** This manuscript is unpublished work carried out during my Ph.D. The authors are Mark A. Currie, Steven P. Smith, and Zongchao Jia. I performed all of the experimental work described in this manuscript. I wrote the initial draft of the manuscript and participated in editing of the manuscript along with Steven Smith and Zongchao Jia.

**Appendix 1:** This paper was published in the Journal of Biological Chemistry in September 2009. The authors are Mark A. Currie, Felipe Merino, Tatiana Skarina, Andrew H.Y. Wong, Alexander Singer, Greg Brown, Alexei Savchenko, Andrés Caniuguir, Victoria Guixé, Alexander F. Yakunin, and Zongchao Jia. Andrew H.Y. Wong solved the apo PhPFK structure. Greg Brown cloned the PhPFK sequence into the vector 11 bacterial expression vectors as well as generated all of the mutants described in the paper. Tatiana Skarina crystallized the apo PhPFK protein and the PhPFK D17A mutant protein bound to AMP. Alexei Savchenko collected the X-ray diffraction data and solved the PhPFK D17A mutant protein structure bound to AMP. Felipe Merino performed the enzymatic assays. I performed structural analysis of both the apo PhPFK and AMP bound D17A mutant PhPFK proteins. Based on this analysis, I modeled the PhPFK protein in its closed conformation and docked F6P and ADP to the active sites. I directed the mutagenesis and enzyme assay experiments based on my structural analysis. I wrote the initial draft of the manuscript and participated in editing of the manuscript along with Victoria Guixé, Alexander F. Yakunin, and Zongchao Jia.
Acknowledgements

I would like to thank my supervisor Dr. Zongchao Jia for his support, guidance, and all of the opportunities that he has given me throughout the course of my study. I would also like to thank Dr. Steven Smith, who has been one of my committee members throughout my graduate studies, for his advice and assistance and for the opportunity to work on the cellulosome project. Thank you to all of the members of the Jia lab both past and present who have contributed to the work described herein or my training through discussions, manuscript editing, and technical assistance: Dr. Melanie Adams-Cioba, Jim Blonde, Dr. Fred Faucher, Neilin Jaffer, Dr. Daniel Lee, Salima Nurmohammed, Dr. Gour Pal, Kateryna Podzelinska, Chelsy Prince, Dr. Mona Rahman, Natalie Roy, Dr. Vinay Singh, Dr. Mike Suits, Laura van Staalduinen, Brent Wathen, Andrew Wong, Dr. Wu Wu, Dr. Susan Yates, Dr. Qilu Ye, Clark Zhang, and Dr. Jimin Zheng. Thank you to the fourth year thesis project students I have supervised: John Li, Jamie (Andrew) Quinlan, Ester Kwok, and Adrienne Cheung; and the volunteers that have assisted me: Chelsy Prince, Monique Mostert, Rob Theiss, and Cassandra Yoon. From Dr. Steven Smith’s lab, I would like to thank Holly Spencer for technical assistance and Dr. Jarrett Adams for useful discussions. Thanks to Dr. Jean-Francois Trempe for teaching me about SAXS and how to collect data, Dr. Carlos Fontes for the plasmids that coexpress the Cel8A enzyme and truncations of the CipA scaffoldin, and Dr. Michal Hammel for assistance with learning the BILBOMD software. I thank the F1 station at Cornell High Energy Synchrotron Source (CHESS, Ithaca, NY), beamline 9-2 at Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, CA), and beamline 19-ID from the Advanced
Photon Source (APS, Argonne, IL) for X-ray diffraction data collection and the SAXS facility at McGill University for SAXS data collection. The work described in this document was funded by operating grants to Dr. Steven Smith and Dr. Zongacho Jia from Canadian Institutes of Health Research. Dr. Jia is a Canada Research Chair in Structural Biology and an NSERC Steacie Fellow. Infrastructure was supported by the Canadian Foundation for Innovation and the Ontario Innovation Trust. Finally, I would like to thank my family for their continuous support and encouragement.
# Table of Contents

Abstract ........................................................................................................................................... ii
Co-Authorship ................................................................................................................................... iv
Acknowledgements .......................................................................................................................... vi
Table of Contents ........................................................................................................................... viii
List of Figures .................................................................................................................................. xii
List of Tables ..................................................................................................................................... xiv
Abbreviations .................................................................................................................................... xv

Chapter 1 General Introduction ...................................................................................................... 1
  1.1 Plant Cell Wall ....................................................................................................................... 1
  1.2 *Clostridium thermocellum* and the Cellulosome ............................................................ 1
  1.3 Cellulosome Ultrastructure ................................................................................................. 2
  1.4 Modular Structure of the Cellulosome ............................................................................... 4
  1.5 The Cohesin-Dockerin Interaction ..................................................................................... 4
  1.6 Lessons from Designer Cellulosomes .............................................................................. 10
  1.7 Reconstructing Protein-Protein Complexes by a Structural “Dissect and Build” Approach ................................................................................................................................. 11
  1.8 X-ray Crystallography ........................................................................................................ 12
  1.9 Small Angle X-ray Scattering ............................................................................................ 15
  1.10 Scope of Thesis .................................................................................................................. 22

Chapter 2 Insights into Higher-Order Organization of the Cellulosome Revealed by a Dissect-and-Build Approach: Crystal Structure of Interacting *Clostridium thermocellum* Multimodular Components ................................................................................................................................. 24
  2.1 Abstract ............................................................................................................................... 24
  2.2 Introduction .......................................................................................................................... 25
  2.3 Materials and Methods, Results, and Discussion .............................................................. 27
    2.3.1 Crystallization and Structural Analysis ....................................................................... 27
    2.3.2 Modular Arrangement within the Complex ............................................................... 29
    2.3.3 Intermolecular Contacts ............................................................................................ 32
2.3.4 Modular Architecture and Flexibility ...............................................................35
2.3.5 Scaffoldin Homodimerization and Higher-Order Cellulosome Organization..36
2.3.6 Dissect-and-Build Strategy for Cellulosome Structure .................................38

Chapter 3 Purification and Crystallization of a Multimodular Heterotrimeric Complex
containing both the Type-I and Type-II Cohesin-Dockerin Interactions from the
Cellulosome of Clostridium thermocellum ............................................................39

3.1 Abstract ........................................................................................................39
3.2 Introduction ..................................................................................................39
3.3 Materials and Methods, Results, and Discussion ............................................41
   3.3.1 Expression, Purification and Complex Formation ....................................41
   3.3.2 Crystallization ......................................................................................43
   3.3.3 Data Collection and Processing ............................................................43

Chapter 4 Structure and Solution Studies of a Multimodular Ternary Protein Complex
from the Cellulosome from Clostridium thermocellum Reveal Dynamics and Higher
Order Organization ................................................................................................47

4.1 Abstract ......................................................................................................47
4.2 Introduction ..................................................................................................48
4.3 Results and Discussion ..................................................................................50
   4.3.1 DocI:CohIγ–X-DocII:CohII Crystal Structure .......................................50
   4.3.2 X-DocII:CohII Region .........................................................................54
   4.3.3 DocI:CohIγ Region ............................................................................55
   4.3.4 Scaffoldin Dimerization ......................................................................56
   4.3.5 Linker Flexibility ...............................................................................57
   4.3.6 SAXS Studies ......................................................................................61
4.4 Materials and Methods ..................................................................................66

Chapter 5 Solution Structure and Dynamics of the Cellulosome from Clostridium
thermocellum Revealed by SAXS ........................................................................69

5.1 Abstract ......................................................................................................69
5.2 Introduction ..................................................................................................70
5.3 Materials and Methods ..................................................................................72
Appendix 1 ADP-Dependent 6-Phosphofructokinase from *Pyrococcus horikoshii* OT3: Structure Determination and Biochemical Characterization of PH1645

A.1.1 Abstract..............................................................................................................147
A.1.2 Introduction .......................................................................................................148
A.1.3 Experimental Procedures ...................................................................................151
  A.1.3.1 Cloning, Expression, and Protein Purification............................................151
  A.1.3.2 Site-Directed Mutagenesis ..........................................................................151
  A.1.3.3 Protein Crystallization ................................................................................151
  A.1.3.4 X-ray Diffraction, Structure Determination, and Refinement ....................153
  A.1.3.5 Determination of Enzyme Activity .............................................................155
  A.1.3.6 Substrate Specificity ...................................................................................155
  A.1.3.7 Kinetic Parameters.......................................................................................155
A.1.4 Results and Discussion ......................................................................................155
  A.1.4.1 Enzymatic Activity and Substrate Specificity of *Ph*PFK ..........................156
  A.1.4.2 Overall Structure .......................................................................................163
  A.1.4.3 Nucleotide and Mg\(^{2+}\) Binding Site.........................................................164
  A.1.4.4 Reaction Mechanism ..................................................................................167
  A.1.4.5 Substrate Induced Fit ..................................................................................167
  A.1.4.6 Sugar Binding Site ......................................................................................170
  A.1.4.7 Differences Between ADP-PFKs and ADP-GKs .......................................173
  A.1.4.8 Conclusion ................................................................................................175
A.1.5 Footnotes/Acknowledgements...........................................................................175
A.1.6 Supplemental Materials .....................................................................................176
  A.1.6.1 Protein Expression and Purification............................................................176
  A.1.6.2 Determination of Enzyme Activity .............................................................177
  A.1.6.3 Hydrodynamic Radius ................................................................................177
  A.1.6.4 Molecular Modeling.....................................................................................179
List of Figures

Figure 1-1 Cartoon and electron micrograph (EM) images of the cellulosome on the surface of *Clostridium thermocellum* in the presence and absence of cellulose ................. 3
Figure 1-2 Modular architecture of the cellulosome from *Clostridium thermocellum* ...... 5
Figure 1-3 Structures of the cohesin and dockerin modules ........................................... 7-8
Figure 1-4 SAXS data, evaluation, and analysis .......................................................... 17-20
Figure 2-1 Structure of the *C. thermocellum* CipA scaffoldin CohI₉–X-DocII trimodular fragment in complex with the SdbA CohII module ..................................................... 31
Figure 2-2 Homodimerization of the *C. thermocellum* CipA scaffoldin CohI₉–X-DocII fragment ....................................................................................................................... 33-34
Figure 3-1 Schematic of the CipA scaffoldin subunit from *C. thermocellum* bound to the type-II cohesin from SdbA and the type-I dockerin from CelD ........................................ 42
Figure 3-2 SDS-PAGE analysis of purified DocI:CohI₉–X-DocII:CohII complex .............. 44
Figure 3-3 Example of DocI:CohI₉–X-DocII:CohII crystals with typical dimensions of approximately 0.25 × 0.25 × 0.20 mm ........................................................................................................ 45
Figure 4-1 Crystal structure of the DocI:CohI₉–X DocII:CohII heterotrimeric complex ...................................................................................................................................... 52-53
Figure 4-2 Interscaffoldin interactions observed in the DocI:CohI₉–X DocII:CohII heterotrimeric complex .................................................................................................................. 59
Figure 4-3 SAXS analysis of the DocI:CohI₉–X DocII:CohII heterotrimeric complex ...... .................................................................................................................................................. 62-64
Figure 5-1 CohI₈—CohI₉-X protein purification and SAXS data ........................................ 74-76
Figure 5-2 CohI₈—CohI₉-X-DocII:CohII protein purification and SAXS data .......... 77-79
Figure 5-3 Cel8A:CohI₈—CohI₉-X-DocII:CohII protein purification and SAXS data ....... .................................................................................................................................................. 80-82
Figure 5-4 Cel8A:CohI₃—CohI₄—CohI₅ protein purification and SAXS data ............. 83-85
Figure 5-5 Cel8A:CohI₂—CBM—CohI₃ protein purification and SAXS data .......... 86-88
Figure 5-6 SAXS based *ab initio* structure of CohI₈—CohI₉–X ........................................ 91-93
Figure 5-7 SAXS based *ab initio* structure of CohI₈—CohI₉–X-DocII:CohII .......... 94-96
List of Tables

Table 2-1 Data collection and refinement statistics ..........................................................30
Table 3-1 Diffraction data and statistics obtained for native DocI:CohI–X-DocII:CohII complex crystals ................................................................................................................46
Table 4-1 Data collection and refinement statistics ..........................................................51
Table 4-2 Contacts between CipA CohI and Cel9D DocI ...............................................58
Table 4-3 Interscaffoldin contacts.....................................................................................60
Table 5-1 Structural parameters from SAXS data ............................................................89
Table A.1-1 Kinetic parameters of wild type and mutant PhPFKs .................................152
Table A.1-2 Crystallographic data and refinement statistics ...........................................154
Table A.1-3 Dynamic light scattering measurements of PhPFK .................................180
Abbreviations

ADP  Adenosine Diphosphate

ADP-GK  ADP-dependent glucokinase

ADP-GK/PFK  Bifunctional ADP-Dependent gluco-Phosphofructokinase from Methanococcus jannaschii

ADP-PFK  ADP-dependent Phosphofructokinase

AMP  Adenosine Monophosphate

APS  Advanced Photon Source

ATP  Adenosine Triphosphate

ATP-GK  ATP-Dependent Glucokinase

ATP-PFK  ATP-Dependent Phosphofructokinase

CBM  Carbohydrate-Binding Module

CDP  Cytidine Diphosphate

CHESS  Cornell High Energy Synchrotron Source

Coh  Cohesin Module

CohI  Type-I Cohesins

CohI_x  The Xth Type-I Cohesin Module from the CipA Scaffoldin from Clostridium thermocellum where X is an integer between 1 and 9

CohII  Type-II Cohesin
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CohIII</td>
<td>Type-III Cohesin Module</td>
</tr>
<tr>
<td>Doc</td>
<td>Dockerin Module</td>
</tr>
<tr>
<td>DocI</td>
<td>Type-I Dockerin Module</td>
</tr>
<tr>
<td>DocII</td>
<td>Type-II Dockerin Module</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EM</td>
<td>Electron Micrograph</td>
</tr>
<tr>
<td>EMBL</td>
<td>European Molecular Biology Laboratory</td>
</tr>
<tr>
<td>EOM</td>
<td>Ensemble Optimization Method</td>
</tr>
<tr>
<td>F6P</td>
<td>Fructose-6-Phosphate</td>
</tr>
<tr>
<td>FBPase</td>
<td>Fructose-1,6 Bisphosphatase</td>
</tr>
<tr>
<td>GDP</td>
<td>Guanosine Diphosphate</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
</tr>
<tr>
<td>IDP</td>
<td>Inosine Diphosphate</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal Titration Calorimetry</td>
</tr>
<tr>
<td>MAD</td>
<td>Multi-Wavelength Anomalous Dispersion</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MES</td>
<td>Minimal Ensemble Search</td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple-Isomorphous Replacement</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>MR</td>
<td>Molecular Replacement</td>
</tr>
<tr>
<td>MS/MS</td>
<td>Tandem Mass Spectrometry</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene Glycol</td>
</tr>
<tr>
<td>PfGK</td>
<td>ADP-GK from <em>Pyrococcus furiosus</em></td>
</tr>
<tr>
<td>PFK</td>
<td>Phosphofructokinase</td>
</tr>
<tr>
<td>PhGK</td>
<td>ADP-GK from <em>Pyrococcus horikoshii</em></td>
</tr>
<tr>
<td>PhPFK</td>
<td>ADP-PFK from <em>Pyrococcus horikoshii OT3</em></td>
</tr>
<tr>
<td>RMSD</td>
<td>Root Mean Squared Deviation</td>
</tr>
<tr>
<td>SAD</td>
<td>Single-Wavelength Anomalous Dispersion</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small Angle X-ray Scattering</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>SIR</td>
<td>Single-Isomorphous Replacement</td>
</tr>
<tr>
<td>SLH</td>
<td>Surface Layer Homology</td>
</tr>
<tr>
<td>TAP</td>
<td>Tandem Affinity Purification</td>
</tr>
<tr>
<td>TiGK</td>
<td>ADP-GK from <em>Thermococcus litoralis</em></td>
</tr>
<tr>
<td>UDP</td>
<td>Uridine Diphosphate</td>
</tr>
<tr>
<td>X</td>
<td>X-Module</td>
</tr>
<tr>
<td>Xyn10A</td>
<td>Xylanase CelX</td>
</tr>
</tbody>
</table>
Chapter 1
General Introduction

1.1 Plant Cell Wall

Plant cells are surrounded by a complex polysaccharide-rich structure, called the plant cell wall, which regulates cell growth and differentiation. The general organization of plant cell wall components and how they change in order to allow cells to grow and expand is still unclear. However, the composition of the cell wall is well established. Primary plant cell walls are composed of interwoven and crosslinked polysaccharides, glycoproteins, esters, minerals, and enzymes. The three major polysaccharides of the plant primary cell wall are cellulose, hemicelluloses, and pectin. Cellulose is a polymer of 1, 4-linked β-D-glucose residues. Hemicelluloses are structurally similar to cellulose. They are branched polysaccharides with 1, 4-linked β-D-hexosyl residue backbone. Pectin is a family of complex polysaccharides that contain the 1, 4-linked α-D-galacturonic acids. Cellulose is the major component of the plant cell wall and makes up about 33% of total plant biomass. It is a hydrophilic, insoluble, crystalline compound that is held together by a network of inter- and intrastrand hydrogen bonds.

1.2 Clostridium thermocellum and the Cellulosome

Clostridium thermocellum is an anaerobic thermophilic bacterium with a rod-like shaped cell body that is found in soil, sediments, and the intestinal tracts of some animals. The primary energy source of C. thermocellum is cellulose. Cellulose is broken down into simple sugars such as glucose, which are then used to generate energy for the bacterium through fermentation (1). The primary byproduct of fermentation is ethanol.
(1). It is this specialized feature that has garnered *C. thermocellum* much attention in biotechnology research (2). With increasing energy consumption, the steady decline of fossil fuels, and growing environmental concerns, development of alternative energy sources such as biofuels is paramount. *C. thermocellum* can convert cellulose, the most abundant renewable natural carbon source, from plant biomass into useable bioethanol (1). The key to the efficient cellulolytic properties of *C. thermocellum* lies in a secreted multienzyme complex called the cellulosome. The cellulosome complexes contain more than 70 different catalytic subunits that are encoded by more than 100 different genes (2). The cellulosomal structure arranges these catalytic subunits in a way that increases the overall catalytic efficiency of the enzymes (3).

**1.3 Cellulosome Ultrastructure**

Electron micrographs of the *C. thermocellum* cell surface both in the presence and absence of a crystalline cellulose substrate have been captured (Figure 1-1) (4-8). These images reveal surprising ultrastructural changes within the cellulosome in response to its substrate. In the absence of cellulose, the cellulosome forms bulbous protuberances on the surface of the cell (Figure 1-1, left) (9). However, when the cellulose substrate is present, the cellulosome forms extended filamentous protrusions that anchor the cell to its substrate (Figure 1-1, right) (9). Additionally, it was noted that cellulosomal structures cluster on the surface of cells in both the resting and active states (9). These structures are referred to as polycellulosomes. In the presence of cellulose, many cellulosome protractions are believed to form channels that direct cellulolytic byproducts towards the cell ensuring their maximal uptake (9;10).
Figure 1-1 Cartoon and electron micrograph (EM) images of the cellulosome on the surface of *Clostridium thermocellum* in the presence and absence of cellulose. EM image of cellulosome in the presence (right) and absence (left) of cellulose (11). Cartoon cellulosomes in the presence and absence of cellulose (center) (11). Figure adapted from (11).
1.4 Modular Structure of the Cellulosome

The cellulosome is typically thought of as beads on a string since it is composed of multimodular proteins with modules connected by linkers of varying lengths. In \textit{C. thermocellum}, the cellulosome is comprised of three main types of multimodular proteins: 1) the non-catalytic scaffoldin subunit CipA, 2) one of three cell surface bound proteins, SdbA, OlpB, and Orf2p, or the anchoring protein Cthe_0736 3) and the catalytic subunits, cellulases, hemicellulases and other hydrolytic enzymes (Figure 1-2) (12). The scaffoldin protein CipA contains a C-terminal type-II dockerin (DocII), X-module of unknown function (X), nine type-I cohesins (CohI), and a cellulose specific carbohydrate-binding module (CBM) (11;13). The catalytic subunits contain a type-I dockerin (DocI) and a catalytic module from the cellulose- or xylanase-degrading families of enzymes, endoglucanases, cellobiohydrolases, β-glucosidases, or xylanases (11;14-16). The cell surface bound proteins SdbA, OlpB, and Orf2p contain one, two, and seven type II cohesin modules (CohII), respectively, along with surface layer homology (SLH) domains that interact with the cell surface (12;17-19). The anchoring protein Cthe_0736 contains seven CohII modules but no SLH domains, which suggest the possibility of non-cell bound cellulosomes in \textit{C. thermocellum} (20).

1.5 The Cohesin-Dockerin Interaction

Cellulosome assembly is mediated by high affinity calcium-dependent interactions between cohesin (Coh) and dockerin (Doc) modules (21;22). To date at least three different types of Coh-Doc modular pairs have been identified, only two of which
Figure 1-2 Modular architecture of the cellulosome from *Clostridium thermocellum*. The enzymatic subunits are depicted in different shades of orange. The CipA scaffoldin contains nine CohI modules (yellow), a CBM (grey), an X-module (pink), and a DocII module (green). The SdbA anchoring protein is shown in blue and the cell surface is shown as a black line. Scaffoldin and anchoring protein inter-modular linkers are shown in stippled yellow and blue, respectively.
are present in *C. thermocellum* (19;23;24). Structurally, Coh modules consist of an elongated nine-stranded β-sandwich in a classical jellyroll topology with a hydrophobic core (Figure 1-3) (25-35). The Doc binding face is formed by an antiparallel β-sheet including strands 8, 3, 6, and 5 (8-3-6-5 face) (25;26;35). The opposite face is formed by a β-sheet comprising strands 9, 1, 2, 7, and 4 (9-1-2-7-4 face) all in an antiparallel arrangement with the exception of strands 1 and 9, which are parallel (25-36). Doc modules are composed of two loop-helix motifs, referred to as F-hands, separated by a linker (Figure 1-3) (25;26;35-37). The two 12-residue loop regions coordinate one Ca\(^{2+}\) ion each in a pentagonal bipyramid configuration typical of EF-hand Ca\(^{2+}\)-binding proteins (25;26;36;37). The CelD DocI and CipA X-DocII from *C. thermocellum* bind calcium with dissociation constants of 0.25 µM and 7µM, respectively (38;39). In the cellulosome of *C. thermocellum*, the type-I Coh-Doc interaction is responsible for sequestering the catalytic subunits on the cellulosomal scaffoldin, while the type-II Coh-Doc interaction anchors the scaffoldin to the cell surface bound proteins (19;23). Therefore, specificity between type-I and type-II Coh-Doc interactions plays a critical role in the assembly of functional cellulosome complexes. Despite the tertiary structural similarities between type-I and type-II Cohs and Docs, they both bind their respective binding partners with great specificity and high affinities, \(K_a\) of 8.0 x 10\(^7\) M\(^{-1}\) and 5.6 x 10\(^8\) M\(^{-1}\) for the type-I and type-II interactions, respectively (25;39;40). Comparison of the crystal structures of the type I and type II Coh-Doc interactions reveals some key difference that account for the distinct specificities (Figure 1-3). The orientation and extent of contact in CohI-DocI and CohII-DocII interactions differ: DocII modules bind
Figure 1-3 Structures of the cohesin and dockerin modules. The CohI (PDB ID: 1ANU) (32), DocI (PDB ID: 1DAQ) (37), CohI:DocI (PDB ID: 1OHZ) (26), and S45A/T46A CohI:DocI (PDB ID: 2CCL) (36) are shown in panels A, B, C, and D, respectively (CohI, yellow; DocI, orange). Panel E and F depict the CohII (PDB ID: 2BM3) (27) and the CohII:DocII-x (PDB ID: 2B59) (25) (CohII, blue; DocII, green; X, pink). Panel G shows the CohIII (PDB ID: 2ZF9) (41) (CohIII, red).
in a parallel/antiparallel arrangement allowing contact with the entire length of both helices of the DocII module; whereas, DocI modules are rotated clockwise by about 20° relative to DocII and make contact through the entire second helix but only the C-terminus of the first helix (25;26). Furthermore, the type-II Coh-Doc binding interface is less charged than the type-I interface (25;26). Another interesting difference between the type-I and type-II interactions is the structural symmetry of DocI modules (25;35;36). This symmetry creates two almost identical CohI binding surfaces on DocI modules allowing them to interact in two different orientations rotated about 180° from one another (Figure 1-3) (35;36). DocII modules are not symmetrical and therefore do not display a dual mode of binding (25). Initially, it was thought that both of the binding interfaces on the DocI were capable of interacting with a CohI module at the same time, which would contribute to the formation of polycellulosome structures. However, biophysical analysis indicates that the CohI and DocI modules interact with a stoichiometry of 1:1 (35;36). Two other hypotheses about the function of DocI binding plasticity have been proposed. The first suggests that two binding modes allow different combinations of enzymes to be generated on the scaffoldin that may not be possible with a single binding mode due to restrictive steric interactions between the bulky catalytic and non-catalytic modules of the cellulosome (35). The second proposes that the position of the enzymes may need to switch, periodically, in order to maximize synergy (12). Since the Coh-Doc interactions are very tight, the second binding interface may facilitate this type of switching (12). The surface of the substrate is constantly changing as the cellulosome acts on it. Switching the position of the enzymes could provide a way for the
complex to maintain the optimal enzyme positions while degrading its substrate (12). However, both of these hypotheses remain to be tested. Furthermore, all of the studies of DocI plasticity have used truncations that only include the DocI modules. What impact an intact enzyme would have on binding mode has yet to be addressed. Therefore, the functional relevance of this binding mode plasticity in cellulosome complexes is unclear.

1.6 Lessons from Designer Cellulosomes

Bulk production of intact homogeneous cellulosome complexes has yet to be achieved due to technical limitations and remains one of the major obstacles in the study of cellulosome function. However, in the absence of intact complexes, several groups have employed recombinant, often chimeric, mini-cellulosome-like complexes, referred to as designer cellulosomes. Homogeneous designer cellulosomes can be produced in bulk and they can be engineered to answer specific questions about that may be more difficult to ascertain in the context of a larger intact complex. Chimeric enzymatic subunits are created by replacing the native DocI module with a DocI of another non-interacting species. Miniscaffolds are then engineered fusing CohI modules from different species, which allows site specific incorporation of chimeric and native enzymes. Two enzymes bound to a miniscaffold display increased activity relative to enzymes free in solution at the same concentration (42). This phenomenon is referred to as the proximity effect. Interestingly, the enzymatic synergy of designer cellulosomes increases with the number of bound enzymes. It has also been shown that incorporation of a substrate targeting CBM module into designer cellulosomes significantly increases the activity of the enzymes (42). Combination of both the proximity and targeting effect
produce the greatest activity (42). Designer cellulosomes have also been used to probe
different enzyme orders and combinations (42). Different non-natural geometries and
binding modes have also been explored using designer cellulosomes along with a
completely covalent cellulosome composed of a single polypeptide chain with a CBM
fused to two different enzymes (43). However, all of these show a reduced activity
relative to designer cellulosomes that mimic more natural cellulosome architecture (43).

1.7 Reconstructing Protein-Protein Complexes by a Structural “Dissect and Build”
Approach

Proteins orchestrate many fundamental cellular activities through specific
interactions with other cellular components. Large scale efforts are underway to
catalogue all protein-protein interactions in a proteome using genetic approaches,
including the yeast two hybrid, and proteomic approaches, such as tandem affinity
purification (TAP) followed by tandem mass spectrometry (MS/MS) to identify the
proteins in the complex. Many soluble and membrane bound or integrated complexes
have been identified that range in size from simple two protein systems to
multicomponent macromolecular machines such as the ribosome, the RNA degradosome,
the protein translocation machinery, the flagella, protein secretory system, and the plant
cell wall degrading cellulosome. Most subunits within complexes that have been
identified interact tightly with one another and in some cases polypeptide chains actually
intertwine with one another preventing their dissociation. However, the limitations of
current technologies for identifying protein-protein interactions make it difficult to detect
transient or low affinity interactions.
One of the biggest challenges for structural biologists is to provide a mechanistic view of these macromolecular machines that carry out fundamental processes of the cell. Ideally, several high resolution structures of these complexes would be solved representing all of the different stages of action. However, practically this is often difficult to achieve due to the large size and intrinsic flexibility of many of these proteins and their complexes. As a result, many structural biologists employ an alternative approach whereby proteins or complexes are divided into subunits, domains, or modules that are more easily studied using high resolution techniques. This is known as the “dissect and build” approach. The aim of this strategy is to ultimately be able to reassemble a functional protein or complex from its component parts. In many cases, this will require the use of more than one structural technique. Moreover, determining the arrangement of the components in a functional protein or complex can also be challenging. In many cases, structural studies of macromolecular complexes require the use of more than one structural technique in order to solve all of the components and then to help arrange them appropriately. Several studies have been published that combine the use of high resolution structures and low resolution structures. These include multimodular proteins and complexes that display flexibility.

1.8 X-ray Crystallography

X-ray crystallography is a fundamental tool for determining atomic level structures of small molecules and macromolecules. It has been an instrumental technique in advancing knowledge as well as elucidating basic laws in many fields including organic and inorganic chemistry, materials science, metallurgy, mineralogy, and biology.
The first structures of biological macromolecules were solved in the late 1950’s beginning with sperm whale myoglobin, an achievement which later garnered the Nobel Prize in Chemistry (44). X-ray crystallography has since been used to solve some 55 000 (early 2010) structures of biological macromolecules including proteins and nucleic acids alone and in complexes. The next highest grossing technique for determining atomic level structures of biological macromolecules is nuclear magnetic resonance or NMR, which has been used to solve a total of ~8 000 (early 2010) additional structures of biological macromolecules.

In order to generate a structure using X-ray crystallography one must first grow crystals of the protein of interest. Several approaches have been developed to screen for crystallization conditions and optimize crystal growth. The general method involves equilibrating a concentrated (typically 5-20 mg/mL for protein) sample of the molecule of interest against an array of screening solutions. Equilibration has been carried out by dialysis, through a capillary, by direct mixing, or by the most common method, vapour diffusion, where a mixture of the screening solution and sample solution are allowed to equilibrate in an air tight chamber containing a bulk reservoir of the screening solution (45). Identifying conditions that will reproducibly grow crystals of the molecule of interest can be difficult or, more commonly, unpredictable (46). Even after obtaining the initial crystallization conditions, further optimization is normally required in order to grow single crystals that are both large enough and able to generate high enough quality diffraction data to solve the structure.
X-ray diffraction studies are usually performed under cryogenic temperatures (~100 K) in order to reduce radiation damage to the crystal (47). Therefore, after obtaining X-ray suitable crystals, an optimal cryo-protectant solution must be identified that will prevent the formation of ice crystals under cryogenic temperatures (47). Ice present during data collection will diffract X-rays and obscure portions of the diffraction from the protein crystal.

For a typical diffraction experiment, crystals are placed in nylon loop and mounted on a goniometer in front of a monochromatic X-ray beam. The crystals are rotated and the diffraction data is recorded by an X-ray detector. The collected images are used to determine the unit cell dimensions, which peak corresponds to which point in reciprocal space, and the space group or symmetry of the crystal. This process is referred to as indexing. Once the data collection is completed, the indexed data is integrated into a single file that incorporates all Miller Indices of the reflections, their intensities, error estimates, and measures of partiality. Individual images are analyzed for spots appearing in more than one image in a process called merging and the relative intensities of spots in each image are scaled. In order to generate an interpretable electron density map in which a model of the structure can be built both the amplitude and phase of the X-rays must be known. The amplitude is measured by the detector. However, all phase information is lost in diffraction experiment. This is referred to as the Phase Problem. Several methods have been developed in order to calculate the initial phase including ab initio or direct methods, Molecular Replacement (MR), Multi- and Single-Wavelength Anomalous Dispersion (MAD and SAD), and Multiple and Single-Isomorphous
Replacement (MIR and SIR) (48-51). A model is built within the electron density map, which is subsequently refined in order to improve the phase and the fit of the model to the diffraction data. The measure of fit between experimental data and the model is called the R-factor.

1.9 Small Angle X-ray Scattering

Small angle X-ray scattering (SAXS) is an emergent tool in the field of structural biology. The first structural studies of biological macromolecules using SAXS were reported in the 1960’s. However, recent advances in instrumentation and analysis methods have improved both the resolution and reliability of SAXS generated structural models and as a result increased the utility of SAXS in structural biology. The structural information that SAXS experiments produce is low resolution, typically 10-50 Å, compared to X-ray crystallography and NMR, which are capable of atomic resolution. However, it does not require the production of diffraction quality crystals for structure determination, like X-ray crystallography, and it is not subject to the inherent size limitations of NMR and electron microscopy. Few proteins that can be expressed and purified yield structures, due in part to the aforementioned limitations of X-ray crystallography, NMR, and electron microscopy. In these cases, SAXS could be an asset to structural genomics initiatives by providing low resolution structural information when high resolution structures cannot be obtained. Even when high resolution structural information is available for part or all of a protein or complex, SAXS provides information about macromolecular folding, unfolding, aggregation, extended
conformations, flexibly linked domains, shape, conformation, and assembly state in solution, which can complement the existing high resolution data.

The experimental setup for SAXS and X-ray crystallography are very similar. Both require a focused, monochromatic beam of X-rays and an X-ray detector in order to measure the intensities of scattered, in the case of SAXS, or diffracted, in the case of crystallography, X-rays. The main difference between these two techniques lies in the sample and how it is mounted in the path of the X-rays. In SAXS, a dilute solution of target molecules is housed in an X-ray permeable chamber or capillary. Since the molecules are floating in solution, X-ray scattering occurs from all possible orientations of the target molecule simultaneously resulting in an averaging effect. For X-ray crystallography, on the other hand, target molecules are crystallized resulting in highly organized symmetrically related molecules within the crystal lattice. Since all of the molecules within the crystal are fixed in one or few orientations, crystals must be rotated during data collection in order to obtain information from other orientations. Although X-ray crystallography yields higher resolution structures, intermolecular contacts required for crystal packing may misrepresent the crystallized molecule by forcing them into non-biologically relevant conformations.

A buffer run or blank is also acquired and subtracted from the protein data, in SAXS experiments, in order to remove the scattering contributed by buffer molecules for A
Raw data

Buffer subtracted

Intensity

$q (\text{Å}^{-1})$

Intensity

$q (\text{Å}^{-1})$

- 7.5 mg/ml
- 5.0 mg/ml
- 2.5 mg/ml
- 1.25 mg/ml
- buffer
No aggregation

$qR_G\ limit = 1.3$

Aggregation

Aggregation that may be ameliorated
\[ q^2 I(q) \]

- **Unfolded**
- **Partially unfolded**
- **Folded**

\[ q (\text{Å}^{-1}) \]
Figure 1-4 SAXS data, evaluation, and analysis. Panel A contains a plot of raw SAXS data collected at several different protein concentrations (top) and a plot of buffer subtracted data for the same protein concentrations (bottom) (52). Panel B shows a Guinier plot for aggregation free (top), aggregated (middle), and aggregated protein that may be removable (bottom) (52). Panel C shows representative Kratky plots for unfolded, partially folded, and folded protein samples (52). Panel D depicts the pair distance distribution function or $P(r)$ for unfolded, multidomain, and globular protein samples (52). Figures from (52).
subsequent structural analysis. The raw data obtained from SAXS analysis is plotted as X-ray scattering intensity (I(q)) versus the scattering angle (q) (Figure 1-4A). General structural features of the target molecule can be determined through alternative representations of the raw data. The Guinier plot, \( \log(I(q)) \) versus \( q^2 \), for instance, yields a straight line from which the radius of gyration (\( R_G \)) and the intensity measured at zero angle \( I(0) \) can be calculated assuming a monodisperse solution of homogenous particles (Figure 1-4B). These can be used to calculate the mass of the particles in solution. The Kratky plot, \( q^2I(q) \) versus \( q \), can be used to measure disorder in a target molecule (Figure 1-4C). Folded domains produce a single parabolic peak, which can be used to estimate molecular mass of the target molecule (Figure 1-4C). Samples containing flexible regions or random coil, on the other hand, yield linear regions at high q values, in the Kratky plot (Figure 1-4C). The pair distribution function or \( P(r) \) can be calculated through Fourier transform of the SAXS curve (Figure 1-4D). \( P(r) \) depicts the probability of finding two electrons at a given distance apart in a target molecule, which is an important tool in SAXS structural modeling. SAXS offers two general tools for structural modeling: *ab initio* modeling, which is constructing a model of your target molecule based on the SAXS data alone; and rigid body modeling, where two or more known components or domains are positioned to best fit the SAXS data. Most *ab initio* SAXS modeling programs employ a genetic algorithm to randomly add, remove, and move dummy atoms until the SAXS curve calculated for the model is no longer improved by subsequent operations. Several *ab initio* models are generated and then averaged. If subsequent independent sets of random operations all yield similar models, there is confidence that
the final model is correct. Rigid body modeling, by contrast, is used when structures of components of a complex or domains of a multidomain molecule are known individually but not in an intact complex or molecule, respectively. In these cases, the individual structures can be used for rigid body modeling. Rigid body modeling programs determine the position of the subunits or domains that best fit the SAXS curve experimentally determined for the complex or intact protein by use of a genetic algorithm or by generating large pools of possible structures from which the best fit to the SAXS data is determined. If any structural information is known prior to initiating either *ab initio* or rigid body modeling, it can be included as restraints during the modeling process.

1.10 Scope of Thesis

Due to the large size and intrinsic flexibility of the cellulosome, structural studies to date have focused mainly on obtaining high-resolution structures of individual modules and the cohesin-dockerin protein-protein interactions. Although these studies have provided a great deal of information about the functional modules of the cellulosome, they have done little to advance our understanding of cellulosomal ultrastructure, modular arrangement, linker dynamics, interscaffoldin interactions, and the underlying synergistic mechanism of the cellulosome.

The studies described herein are aimed at expanding our structural knowledge from individual modules of the cellulosome to include larger multimodular fragments of the cellulosome. This includes the two largest multimodular high resolution structures of the cellulosome described to date, the CohI$_y$—X-DocII:CohII and the DocI:CohI$_y$—X-
Interscaffoldin contacts were observed in both crystal structures and were confirmed to exist in solution, which has intriguing implications in the formation of polycellulosomal structures \textit{in vivo}. The B-factors of the CohI$_9$—X linker are elevated in both structures indicating a degree of flexibility within this region. In fact, structural alignment of the heterodimeric complex and the four ternary complexes solved from the asymmetric unit reveal five distinct conformations of the CohI$_9$—X linker. SAXS analysis of the ternary complex suggests limited flexibility of the linker in solution. In addition, the ternary crystal structure provides the first evidence for a type-I cohesin-dockerin binding orientation bias. Finally, SAXS studies including ab initio modeling, rigid body modeling, and solution dynamics studies were carried out on several larger cellulosomal complexes including type-I dockerin containing Cel8A enzymes. These SAXS structural studies are providing a comprehensive framework for building the first model of an entire cellulosome.
Chapter 2

Insights into Higher-Order Organization of the Cellulosome Revealed by a Dissect-and-Build Approach: Crystal Structure of Interacting Clostridium thermocellum Multimodular Components

2.1 Abstract

Cellulosomes are large, multienzyme, plant cell wall-degrading protein complexes found affixed to the surface of a variety of anaerobic microbes. The core of the cellulosome is a noncatalytic scaffoldin protein, which contains several type-I cohesin modules that bind type-I dockerin-containing enzymatic subunits, a cellulose-binding module, an X-module, and a type-II dockerin that interacts with type-II cohesin-containing cell surface proteins. The unique arrangement of the enzymatic subunits in the cellulosome complex, made possible by the scaffoldin subunit, promotes enhanced substrate degradation relative to the enzymes free in solution. Despite representative high-resolution structures of all of the individual modules of the cellulosome, this mechanism of enzymatic synergy remains poorly understood. Consequently, a model of the entire cellulosome and a detailed picture of intermodular contacts will provide more detailed insight into cellulosome activity. Toward this goal, we have solved the structure of a multimodular heterodimeric complex from Clostridium thermocellum composed of the type-II cohesin module of the cell surface protein SdbA bound to a trimodular C-terminal fragment of the scaffoldin subunit CipA to a resolution of 1.95 Å. The linker that connects the ninth type-I cohesin module and the X-module has elevated temperature factors, reflecting an inherent flexibility within this region. Interestingly, a novel dimer
interface was observed between CipA and a second, symmetry-related CipA molecule within the crystal structure, mediated by contacts between a type-I cohesin and an X-module of a symmetry mate, resulting in two intertwined scaffoldins. Sedimentation velocity experiments confirmed that dimerization also occurs in solution. These observations support the intriguing possibility that individual cellulosomes can associate with one another via inter-scaffoldin interactions, which may play a role in the mechanism of action of the complex.

2.2 Introduction

Microbial degradation of the plant cell wall necessitates an extensive array of hydrolytic enzymes to degrade such a heterogeneous, insoluble, and highly recalcitrant substrate (53). Indeed, the physical and chemical complexity of the polysaccharides within the plant cell wall demands that these enzymes act in synergy to degrade this composite structure. In order to efficiently degrade plant cell wall polymers in a synergistic manner, some anaerobic bacteria have developed a supramolecular multienzyme complex, termed the cellulosome (11;54-56).

The cellulosome of Clostridium thermocellum has been the most extensively studied from both a biochemical and structural perspective. The central cellulosomal subunit is the multimodular, noncatalytic scaffoldin protein that organizes the various enzymes into a cohesive functional entity. It comprises a cellulose-specific carbohydrate-binding module, which targets the enzyme complex to its substrate, nine type-I cohesin (CohI) modules, an X-module, and a C-terminal type-II dockerin (DocII) module (13;14).
The CohI modules mediate the integration of the catalytically active subunits into the complex via a high-affinity, noncovalent interaction with the enzyme-associated type-I dockerin (DocI) modules (22;26;36), whereas the DocII module associates the cellulosome complex to the bacterial cell surface via binding to the CohII modules of the cell surface anchoring subunits SdbA, Orf2p, and OlpB (19;57).

Substantial efforts have been undertaken to elucidate the tertiary and quaternary structural features of the cellulosome. Electron microscopy studies indicated that polycellulosome organelles are located on the cell surface and appear as extended protuberances in the presence of a cellulose substrate (7;9). Small-angle X-ray scattering studies have more recently illustrated that the conformational flexibility provided by the linker regions between the scaffoldin CohI modules allow for optimal positioning of the enzymatic subunits onto the substrate (58;59). However, the major focus has been in obtaining atomic resolution structural information of the cellulosome components through a ‘piecemeal’ approach, whereby representative isolated modules of the scaffoldin, catalytic and cell-surface subunits have been the targets of structure determination by X-ray crystallography and NMR spectroscopy (60). These include several C. thermocellum catalytic modules (61-67), the CipA carbohydrate-binding module (68), a DocI module (37), the second and seventh CohI modules (32;34), and the SdbA CohII module (27). Recent structures of the type-I (26;36) and type-II (25) Coh–Doc complexes have extended this structural knowledge toward the molecular determinants driving cellulosome assembly. This latter structure comprised the C-terminal X-module–DocII (X-DocII) modular pair of the CipA scaffoldin and represented
a shift from the piecemeal approach toward a ‘dissect-and-build’ concept, which involves structure determination of multimodular fragments of cellulosomal subunits. An unexpected extensive modular interface between the DocII module and the neighbouring X-module was identified, which reveals that the C-terminal region of the CipA scaffoldin subunit has an elongated topology (25). Despite these advances there is little detailed structural information on the intact cellulosome.

In this study, we extend the dissect-and-build concept to determine the X-ray crystal structure of the largest multimodular portion of the C. thermocellum cellulosome complex to date, which contains the C-terminal trimodular fragment of the CipA scaffoldin (the ninth CohI, connected by a linker to the X-DocII) bound to the SdbA CohII module. The structure reveals an elongated topology with a flexible 13-residue linker connecting the ninth CohI module (CohI₉) and the X module. An intermolecular interface, composed of van der Waals and hydrogen-bonding contacts, is observed between strands 4 and 5 of the CohI₉ module and two loops of an X-module from a neighbouring scaffoldin molecule. These studies provide fundamental structural insights into cellulosome modular architecture and intermolecular scaffoldin interactions.

2.3 Materials and Methods, Results, and Discussion

2.3.1 Crystallization and Structural Analysis

A C-terminal trimodular fragment consisting of the CohI₉, X-module, and type-II dockerin module (CohI₉–X-DocII) of the C. thermocellum CipA scaffoldin, which contains a single point mutation (Cys269Ser) to prevent disulfide-mediated self-
association, and the SdbA type-II cohesin (CohII) were expressed as N-terminal His-
tagged fusion proteins. Expression and purification of recombinant CohII was performed as described previously (69). The CipA CohI9–X-DocII construct was expressed and purified from inclusion bodies in a manner similar to that described for the construct of the \textit{C. thermocellum} CipA X-DocII modular pair (39). The CohI9–X-DocII:CohII heterodimeric complex was generated by combining purified CohI9–X-DocII and CohII in a 1.3:1 molar ratio followed by application to a HiLoad16/60 Superdex 75 size-
exclusion column in 25 mM Hapes (pH 7.5), 50 mM NaCl, and 5 mM CaCl$_2$. The purified CohI9–X-DocII:CohII complex was pooled, concentrated using a Millipore Amicon 10-kDa centrifugal device, and crystallized by the hanging-drop vapour-
diffusion method at 21 °C. The CohI9–X-DocII:CohII complex (25 mg/mL) and reservoir solution [100 mM Mes (pH 6.5), 200 mM ammonium sulfate, 30\% (w/v) PEG (polyethylene glycol) 3350] mixed in a 1:1 ratio (4 μL) were equilibrated against 1 mL of reservoir solution. Crystals formed approximately 7 days after plating. Reservoir solution supplemented with 15\% PEG 400 was used as a cryoprotectant for the crystals prior to flash-cooling in a cryostream of N$_2$ gas. An X-ray diffraction data set was collected at 100 K at the F1 station of CHESS (Cornell High Energy Synchrotron Source, Ithaca, NY).

The data set was indexed and scaled using HKL2000 (70). The CohI9–X-
DocII:CohII complex was crystallized in a P2$_1$2$_1$2$_1$ space group (Table 2-1). Phasing was performed using molecular replacement (PHASER (71)) and structures of the X-
DocII:CohII complex (25) and the second CohI module from CipA as the probes (32).
The structures were built with ARP (72) and refined to 1.95 Å resolution with REFMAC (73) using cycles of wARP for solvent positions (72). A single copy of the heterodimeric complex was found in the asymmetric unit. Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession number 3KCP. The structure included residues 32–194 from the CohII module, residues 1548–1850 from the CohI–X-DocII trimodular fragment of the CipA scaffoldin subunit, 244 water molecules, two calcium ions, and a single chloride ion (Figure 2-1).

2.3.2 Modular Arrangement within the Complex

The heterodimeric CohI–X-DocII:CohII complex can be divided into two distinct globular regions (Figure 2-1). One region comprises CohII and X-DocII modular pair despite belonging to two independent polypeptide chains, while CohI constitutes the second globular moiety covalently linked to X-DocII. The elongated topology of the X-DocII:CohII is similar to that of the X-DocII:CohII complex described previously (25) (backbone r.m.s.d. of 0.35 Å), with the structures of the individual modules and the X-DocII and DocII–CohII intermodular interfaces maintained. This region is tethered to CohI, which displays the characteristic nine-stranded β-jellyroll topology (Figure 2-1, yellow), by an extended 13-residue linker (Val1687–Lys1699) that places the X module and the CohI module at a distance of 18 Å from each other. The temperature factors associated with the linker are elevated when compared to those of residues in the protein module cores, indicating that this region is dynamic and could allow the CohI module to explore conformational space, including coming into closer proximity with the X-DocII:CohII region.
Table 2-1 Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>CohIg–X-DocIi:CohIi</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space Group</strong></td>
<td>$P2_12_12$</td>
</tr>
<tr>
<td><strong>Cell dimensions</strong></td>
<td></td>
</tr>
<tr>
<td>$a$, $b$, $c$ (Å)</td>
<td>56.14, 58.32, 170.53</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.912</td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>50.0-1.95 (2.05-1.95)</td>
</tr>
<tr>
<td><strong>$R_{sym}$ (%)</strong></td>
<td>6.6 (42.4)</td>
</tr>
<tr>
<td><strong>$I/\sigma I$</strong></td>
<td>36.8 (3.2)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>99.4 (98.7)</td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>6.8 (5.5)</td>
</tr>
</tbody>
</table>

**Refinement**

| **Resolution (Å)** | 42.49-1.95 |
| **No. reflections** | 39350      |
| **$R_{work}/R_{free}$ (%)** | 20.7/23.1 |
| **$R_{p.i.m.}$ (%)$^b$** | 2.7        |
| **$R_{meas/r.i.m.}$ (%)$^c$** | 7.2        |
| **No. of atoms** |                        |
| Protein | 3542       |
| Ions | 3         |
| Water | 244        |

**$B$-factors (Å$^2$)**

<table>
<thead>
<tr>
<th><strong>B-factors (Å$^2$)</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>35.75</td>
</tr>
<tr>
<td>Ions</td>
<td>39.15</td>
</tr>
<tr>
<td>Water</td>
<td>48.81</td>
</tr>
</tbody>
</table>

**Bond lengths (Å)**

<table>
<thead>
<tr>
<th><strong>Bond lengths (Å)</strong></th>
<th>0.019</th>
</tr>
</thead>
</table>

**Bond angles (°)**

<table>
<thead>
<tr>
<th><strong>Bond angles (°)</strong></th>
<th>1.679</th>
</tr>
</thead>
</table>

**Ramachandran plot values (%)**

<table>
<thead>
<tr>
<th><strong>Ramachandran plot values (%)</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in most favoured regions</td>
<td>88.9</td>
</tr>
<tr>
<td>Residues in additionally allowed regions</td>
<td>11.1</td>
</tr>
</tbody>
</table>

$^a$ Values in parentheses indicate the statistics for the highest resolution shell.

$^b$ $R_{p.i.m.} = \frac{\sum_{h}(1/n_h - 1)\sum_{l}|I_{hl} - \langle I_h \rangle|/\sum_h \sum_l |I_{hl}|}{\sum_h \sum_l |I_{hl}|}$, where $n_h$ is the number of observations of reflection $h$, $I_{hl}$ is the $l$th observation, and $\langle I_h \rangle$ is the weighted average intensity for all observations $l$ of reflection $h$ (74).

$^c$ $R_{meas/r.i.m.} = \frac{\sum_{h}(n_h/n_h - 1)\sum_{l}|I_{hl} - \langle I_h \rangle|/\sum_h \sum_l |I_{hl}|}{\sum_h \sum_l |I_{hl}|}$, where $n_h$ is the number of observations of reflection $h$, $I_{hl}$ is the $l$th observation, and $\langle I_h \rangle$ is the weighted average intensity for all observations $l$ of reflection $h$ (74).
Figure 2-1 Structure of the *C. thermocellum* CipA scaffoldin CohI₉–X-DocII trimodular fragment in complex with the SdbA CohII module. The backbone ribbon representation of the complex depicts SdbA CohII in blue, DocII in green, X module in rose, and CohI₉ in yellow. The calcium ions and chloride ion appear as orange and cyan spheres, respectively. The modules are identified, and the N- and C-termini are labelled accordingly.
2.3.3 Intermolecular Contacts

Homodimerization of the CohI<sub>9</sub>–X-DocII:CohII complex (i.e., homodimerization of the heterodimeric complex) is observed in the crystal lattice, where an intermodular interface occurs between strands 4 and 5 of a CohI<sub>9</sub> module from one molecule and the N-terminus and loop region connecting strands 4 to 5 of the X-module from an adjacent molecule (Figure 2-2a). The interface involves residues from CohI<sub>9</sub> and the X-module forming van der Waals contacts (CohI<sub>9</sub>: Glu1593, Ile1595, Glu1596, Ala1615, Val1616, Tyr1717, Pro1618; X module: Asn1698, Glu1745, Tyr1747, Tyr1781, Val1783, Ala1797, Ala1799) and hydrogen bonds (CohI<sub>9</sub>: Ile1595, Glu1596, Glu1598, Val1616, Asp1619, Asp1674; X-module: Glu1696, Gly1697, Tyr1734, Ala1748, Ala1750, Ser1751).

Homodimerization produces symmetrical contacts between both CohI<sub>9</sub> modules and the X-module of the other molecule such that the two molecules become intertwined (Figure 2-2b). The intermolecular interface encompasses an occluded surface area of 2008 Å².

Notably, despite this extensive intermolecular interface, the 8-3-6-5 face of the CohI<sub>9</sub> module remains exposed for binding to the DocI module.

To assess the oligomeric properties of the CohI<sub>9</sub>–X-DocII:CohII complex in solution, sedimentation velocity experiments were performed. Sedimentation profiles were evaluated by a continuous distribution c(S) Lamm equation model.

Sedimentation coefficient distribution for a CohI<sub>9</sub>–X-DocII:CohII sample at a concentration of 100 µM revealed three peaks at 1.57, 3.02, and 4.24 S. These correspond to molecular masses of 24.7, 52.2, and 112 kDa, respectively, and represent three potential modular arrangements: the isolated CohII module (20.5 kDa), the
Figure 2-2 Homodimerization of the *C. thermocellum* CipA scaffoldin CohI₉–X-DocII fragment. (a) The backbone ribbon representation of the complex in a translucent surface illustrating the CohI₉ module of a symmetry molecule (grey) at the CohI₉–X intermodular cleft of the CohI₉–X-DocII:CohII fragment. Enlargement of the boxed region illustrates the residues at the intermolecular CohI₉–X dimer interface, depicted as sticks, with the residues identified by their one-letter amino acid code and their number in sequence. The CohI₉ from the symmetry molecule is depicted in grey and the X module in rose. (b) Backbone ribbon representation with a translucent surface of the CohI₉–X-DocII:CohII homodimer, illustrating the intertwining of the two fragments. The modular components of one of the heterodimers (CohI₉–X-DocII:CohII) are coloured as in (a) and its symmetry-related mate is coloured grey. The inset depicts the ribbon representation of the two CohI₉ modules in the homodimer, where the 8-3-6-5 face (labelled accordingly) of each CohI₉ is exposed and accessible for binding of a dockerin module. Calcium ions are depicted as orange spheres and the chloride ion as a cyan sphere.
monomeric form of the CohI\textsubscript{9}–X-DocII:CohII heterodimer (55 kDa), and a homodimeric species of the CohI\textsubscript{9}–X-DocII:CohII complex (110 kDa). At this concentration, the latter species represents 3.5% of the total signal, while at 30 µM the homodimeric species represents only 1% of the total signal. Thus, the CohI\textsubscript{9}–X-DocII:CohII complex can homodimerize in solution and is not simply an artifact of the crystallization process.

2.3.4 Modular Architecture and Flexibility

The modular structure of the \textit{C. thermocellum} CipA scaffoldin has historically been depicted as ‘beads-on-a-string’, with the protein modules separated by unstructured and flexible linker regions. These linkers contain glycosylated Thr/Pro-rich sequences of approximately 20 residues (75), which are similar to the highly flexible linker region between the catalytic and cellulose-binding modules of xylanase CelX (Xyn10A) from \textit{Cellulomonas fimi} (76). Electron microscopy studies of intact cellulosomes (7;9) and more recent small-angle X-ray scattering studies on recombinantly produced cellulosome-like complexes (58;59) also support such a flexible model. As such, it has been suggested that a precise crystal structure of the entire cellulosome is unlikely due, in part, to the inherent structural flexibility of the scaffoldin (77). However, our unexpected observation of intermodular interface between the DocII and X-module at the C-terminus of the \textit{C. thermocellum} CipA scaffoldin (25) deviated from this dogma and presented the possibility that additional intermodular contacts may exist. Furthermore, recent structural studies of adjacent cohesin modules from the cellulosome of \textit{Acetivibrio cellulolyticus} illustrated that crystallization of cellulosomal modules with intact linker segments is
possible, and that the crystal packing and linker–module interactions provide novel insights in linker dynamics, cellulosome organization, and ultrastructure (30).

The current findings indicate that the X-module and the adjacent CohI₀ are separated by a 13-residue linker and do not contact one another in the CohI₀–X-DocII:CohII complex structure. While the entire linker region in the CohI₀–X-DocII:CohII heterodimeric complex structure was modelled from electron density, it appears to have a degree of conformational freedom based on the elevated temperature factors. However, an intermolecular interaction between the X-module of one complex molecule and the CohI₀ module from a symmetry-related molecule appears to limit the extent of the conformational space explored by this linker. This interaction likely explains, in part, why the entire linker was defined by density in the CohI₀–X-DocII:CohII complex and why crystallization of the complex was possible.

2.3.5 Scaffoldin Homodimerization and Higher-Order Cellulosome Organization

The intermolecular CohI₀–X module association leads to the homodimerization of the binary complex. Crystallographic oligomers have been previously observed for cellulosomal constructs including the isolated CipA CohI modules, where the interface was localized to the 8-3-6-5 face of each molecule and overlapped with the DocI-binding site, and the DocI:CohI complex, which formed crystallographic dimers and trimers, respectively (26;32;34). While the potential for a similar homodimeric interface existed for our CohI₀–X-DocII:CohII complex, no such observation was made. Rather, loop regions of the X-module and strands 4 and 5 of CohI₀ from a symmetry-related molecule
form the intermolecular interface. The concentration-dependent homodimerization of the CohI₀–X-DocII:CohII complex in solution suggests that this phenomenon cannot be entirely attributed to crystallization and may have some \textit{in vivo} relevance. Some cell-surface subunits, such as Orf2p and OlpB, comprise multiple CohII modules and thus have the ability to bind multiple scaffoldin subunits (18;19;78). In such an environment, the local concentration of the scaffoldin subunits near the bacterial cell surface may be sufficient to induce homodimerization, which may contribute to the stability of higher-order cellulosome arrangements.

The notion of polycellulosomes was first proposed based on electron microscopic studies, where cellulosome clusters were observed on the bacterial cell surface (7;9). However, the mechanisms by which these structures form and are stabilized is not well understood. The complex structure presented here provides an intriguing potential mechanism for stabilizing, at least in part, cellulosome ternary structure. The homodimerization of this complex involves two intermolecular X-module–CohI₀ interactions, where the X-module from one molecule contacts the CohI₀ from a symmetry-related molecule. These contacts result in a twisting of the two molecules around one another, yet they do not block the DocI binding surface on the CohI₀ module (Figure 2-2b, inset). The intrinsic flexibility of the linker between the X-module and the CohI₀ in each CipA monomer is thus able to accommodate conformational changes while maintaining the intermolecular contacts.
2.3.6 Dissect-and-Build Strategy for Cellulosome Structure

The use of a ‘dissection’ concept in the structural characterization of the cellulosome has allowed for a nearly complete complement of representative structures of the various isolated modules associated with the *C. thermocellum* cellulosome. The recent crystal structures of the *C. thermocellum* DocI:CohI and X-DocII:CohII complexes have extended this approach toward ‘building’ the cellulosome (25;26;36). While these structures have advanced our understanding of cellulosome assembly and function, a comprehensive picture of overall cellulosome modular architecture and polycellulosome formation has yet to emerge. The dissect-and-build approach and resultant findings of this study extend our understanding of CipA scaffoldin structure and provide novel insights into higher-order scaffoldin interactions.
Chapter 3
Purification and Crystallization of a Multimodular Heterotrimeric Complex containing both the Type-I and Type-II Cohesin-Dockerin Interactions from the Cellulosome of Clostridium thermocellum

3.1 Abstract
The central component of the cellulosome, a multienzyme plant cell wall degrading complex, from Clostridium thermocellum is the multimodular scaffoldin subunit CipA. It captures secreted cellulases and hemicellulases and anchors the entire complex to the cell surface via high-affinity, calcium-dependent interactions between cohesin and dockerin modules termed type-I and type-II, respectively. We report the crystallization of a heterotrimeric complex comprising the type-II cohesin module from cell surface protein SdbA, a trimodular C-terminal fragment of the scaffoldin CipA, and the type-I dockerin module from the CelD cellulase. The crystals belong to space group P2_12_12_1, with unit-cell parameters a = 119.37, b = 186.31, c = 191.17 Å. Crystals diffracted to 2.7 Å with four or eight molecules of the ternary protein complex in the asymmetric unit.

3.2 Introduction
The cellulosome is a large cell-surface bound multienzyme complex responsible for the degradation of cellulose and other plant cell wall polysaccharides. Originally discovered in the thermophilic anaerobe Clostridium thermocellum (79;80), cellulosomes have since been identified in a variety of other anaerobic bacteria, ruminal
bacteria, and anaerobic fungi (for reviews, see 14;54;55;81;82). The cellulosome of *C. thermocellum* is composed of three modular protein components: cellulases, hemicellulases and other hydrolytic enzymes, the CipA scaffoldin subunit (13), and one of three cell surface associated proteins (SdbA, OlpB, Orf2p; (19;57;78)(Figure 3-1). Cellulosome assembly is mediated by two types of high affinity calcium-dependent interactions between cohesin (Coh) and dockerin (Doc) modules. The type-I interaction is responsible for localizing the various enzymes to the scaffoldin while the type-II interaction anchors the scaffoldin to a cell surface associated protein. The CipA scaffoldin contains nine type-I Coh modules, a type-II Doc module, an X-module of unknown function, and a cellulose-specific carbohydrate-binding module (CBM), all of which are connected by flexible linkers of varying length (Figure 3-1).

In order to obtain a better structural understanding of the scaffoldin, crystal structures have been obtained for the individual type-I and type-II Coh and Doc modules (32;34;37) and their complexes (25;26;36), the CBM (63), and the X-module (25). However, few multimodular structures have been determined, likely due to the inherent flexibility of the linker regions that separate each module. Here, we report the generation, purification, crystallization, and preliminary X-ray characterization of a heterotrimeric multimodular complex including the three C-terminal modules of CipA (residues 1533-1853) bound to the type-I Doc module (DocI) from the CelD cellulase (residues 549-625) and the type-II Coh (CohII) module from the cell surface protein SdbA (residues 27-200).
3.3 Materials and Methods, Results, and Discussion

3.3.1 Expression, Purification and Complex Formation

The C-terminal CohI\textsubscript{9}–X-DocII fragment of CipA with a C-terminal hexahistidine tag, SdbA CohII with an N-terminal hexahistidine tag, and CelD DocI with an N-terminal dodecahistidine tag were all recombinantly expressed and purified in a similar manner. Briefly, transformed BL21 (DE3) cells (Novagen) were grown in LB medium supplemented with 100 mg/L of ampicillin at 37 °C while shaking until an $\text{OD}_{600}$ of 0.6. IPTG was added to a final concentration of 1 mM and growth was continued for an additional 4 h. Cells were harvested by centrifugation (20 min at 3000g), resuspended in 20 mL of buffer A (25 mM Tris–HCl pH 7.4, 250 mM NaCl, 8 M urea), and lysed on ice by sonication. The insoluble fraction was removed by centrifuging at 20000g in a Beckman JA-20 rotor for 20 min. The supernatant was applied to a Ni\textsuperscript{2+}-charged chelating column that was pre-equilibrated in buffer A. The column was subsequently washed with buffer A containing 20 mM imidazole and the bound protein was eluted with buffer A containing 400 mM imidazole. Purified CohI\textsubscript{9}–X-DocII, DocI, and CohII, were pooled and refolded by dialysis into buffer B (20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM CaCl\textsubscript{2}, and 1 mM DTT). The refolded complex was separated from excess unbound proteins by size exclusion chromatography using a Hi-load 16/60 Superdex 200 size-exclusion column (Amersham Pharmacia Biosciences) equilibrated in buffer B. The protein was eluted using buffer B in 4 mL fractions and the purity was confirmed via SDS-PAGE with Coomassie Blue staining (Figure 3-2).
Figure 3-1 Schematic of the CipA scaffoldin subunit from *C. thermocellum* bound to the type-II cohesin from SdbA and the type-I dockerin from CelD. Cohesins, dockerins, the X-module, and the carbohydrate-binding module are labelled, C, D, X, and CBM, respectively. Type-I cohesins and dockerins are shown in grey and white, respectively. Type-II cohesins and dockerins are shown in black and grey, respectively. The complex described in the current article is surrounded by a box.
3.3.2 Crystallization

The DocI:CohI–X-DocII:CohII complex crystals were grown by hanging drop vapour diffusion with a drop containing 2 µL protein at 28 mg/mL, 2-4µL of reservoir solution of 100 mM HEPES pH 7-7.75, 1.3-1.5 M lithium sulphate, and 0.5 µL of 1 M potassium sodium tartrate. Crystals took 7-10 days to grow at room temperature. The crystals were tetragonal in shape with measurements of 0.25 × 0.25 × 0.20 mm (Figure 3-3).

3.3.3 Data Collection and Processing

X-ray data were collected at beamline 9-2 at the Stanford Synchrotron Radiation Lightsource using a MarUSA MarMosaic 325 CCD detector. Data was collected at 100 K with crystals that were soaked in reservoir solution containing 20% glycerol as a cryo protectant and flash frozen in liquid nitrogen.

The crystals were of a primitive orthorhombic space group P2₁2₁2₁, with unit-cell dimensions of a = 119.37, b = 186.31, c = 191.17 Å. A Matthews coefficient of 4.08 Å³ Da-1 and 2.04 Å³ Da-1 were obtained with solvent contents of 69.85% and 39.70% for an asymmetric unit containing four and eight heterotrimeric protein complexes, respectively. The data was processed to 2.7 Å resolution with an Rsym of 4.6% (Table 3-1) using HKL2000 (70). A molecular replacement strategy based on the X-Doc:CohII structure (25) and the CohI:DocI structures (26;36), is being employed to solve the DocI:CohI–X-DocII:CohII heterotrimeric complex structure.
Figure 3-2 SDS-PAGE analysis of purified DocI:CohI$_9$–X-DocII:CohII complex. Lane 1, molecular weight marker; lane 2, purified DocI:CohI$_9$–X-DocII:CohII complex.
Figure 3-3 Example of Doc1:CohI9–X-DocII:CohII crystals with typical dimensions of approximately $0.25 \times 0.25 \times 0.20$ mm.
Table 3-1 Diffraction data and statistics obtained for native DocI:CohI–X-DocII:CohII complex crystals.

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>$P2_12_12_1$</td>
</tr>
<tr>
<td>Unit Cell Parameters (Å)</td>
<td>$a = 119.37, b = 186.31, c = 191.17$</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97927</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
</tr>
<tr>
<td>Resolution Range (Å)</td>
<td>30.0 - 2.7 (2.8-2.7)</td>
</tr>
<tr>
<td>Observed Reflections</td>
<td>113244</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>107440</td>
</tr>
<tr>
<td>Data Completeness (%)</td>
<td>95.8 (78.5)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.6 (2.4)</td>
</tr>
<tr>
<td>$R_{sym}$ (%)</td>
<td>4.6 (75.5)</td>
</tr>
<tr>
<td>$&lt;I/σI&gt;$</td>
<td>30.9 (1.9)</td>
</tr>
<tr>
<td>Matthews coefficient (Å³ Da⁻¹)</td>
<td>4.08/2.04*</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>69.85</td>
</tr>
</tbody>
</table>

$R_{sym} = \frac{\sum |I(k) - \langle I \rangle|}{\sum I(k)}$ where $I(k)$ and $I$ represent the diffraction intensity values of the individual measurements and the corresponding mean values. Values in parentheses are for the highest resolution shell.

*four/eight heterotrimeric protein complexes per asymmetric unit.
Chapter 4

Structure and Solution Studies of a Multimodular Ternary Protein Complex from the Cellulosome from Clostridium thermocellum Reveal Dynamics and Higher Order Organization

4.1 Abstract

Cellulosome complexes efficiently degrade plant cell wall polysaccharides by the enhanced cellulolytic activity of hydrolytic enzymes conferred by the binding of a core non-enzymatic cellulosomal scaffoldin subunit. However, understanding the unique quaternary structural elements responsible for the enzymatic synergy of the cellulosome is hindered by both the large size and inherent flexibility of these multi-protein complexes. Herein, we have used X-ray crystallography and small angle X-ray scattering (SAXS) to structurally characterize a ternary protein complex of the CipA scaffoldin subunit comprising the type-II dockerin module, X-module, and the ninth type-I cohesin module from Clostridium thermocellum bound to the type-II cohesin from the SdbA anchoring protein, and the type-I dockerin from the family 9 cellulosomal glycoside hydrolase, Cel9D. The heterotrimeric structure containing five modules was solved to 2.7 Å with four complexes in the asymmetric unit. Both the solution structure and the crystal structure of the complex reveal two rigid domains composed of the type-I cohesin-dockerin interaction and the type-II cohesin-dockerin-X-module separated by a short 13-residue linker in an extended conformation. Surprisingly, all four type-I dockerins in our crystal structure bind to type-I cohesins in the opposite orientation compared to previous wild-type type-I cohesin-dockerin structures. This is the first evidence of orientation bias.
for type-I dockerins in the classical cellulosome system. Contacts, between X-modules and symmetry-related type-I cohesins, were observed in all four complexes. In solution, inter-scaffoldin dimerization occurs only in the absence of the type-I dockerin. SAXS analysis of the complex indicates that the linker is highly flexible; however, the direction of flexibility is limited to a single plane. This work provides insight into the behaviour of the linker between the X-module and the last type-I cohesin in solution along with inter-scaffoldin interactions, suggesting restrained and highly selective flexibility for optimal cellulosome function. Furthermore, the structure determined here represents the largest fragment of the cellulosome to be solved to date and the only fragment to contain three different proteins.

### 4.2 Introduction

Plant cell wall polysaccharides are the most abundant renewable carbon source on earth. However, the composite heterogeneous structure of the plant cell wall makes it a recalcitrant substrate and an obstacle to exploiting this rich carbon source (53). Some anaerobic microbes have developed a specialized apparatus, dubbed cellulosomes, capable of efficient degradation of plant cell walls through the synergistic activity of various secreted cellulases and hemicellulases (11;54-56).

The prototypical cellulosome from the thermophilic anaerobe *Clostridium thermocellum* was the first to be discovered and remains the best studied. The core of this cellulosome is composed of a multimodular non-catalytic scaffold protein, CipA, that binds and optimally positions secreted cellulolytic enzymes while remaining firmly
anchored to cell surface bound proteins. CipA contains a cellulose-specific carbohydrate-binding module (CBM), which targets the enzyme complex to its substrate, nine type-I cohesin (CohI) modules, an X-module and a C-terminal type-II dockerin (DocII) module (13;14). Cellulosome assembly is mediated by high-affinity calcium-dependent interactions between cohesin and dockerin modules. CipA attaches to the CohII modules of cell surface-anchored proteins, SdbA, Orf2p, and OlpB via its DocII module and captures secreted DocI-containing cellulolytic enzymes on its CohI modules (19;22;26;36;57).

Cellulolytic enzymes bound to the cellulosomal scaffold display enhanced activity relative to enzymes free in solution (42;83). An understanding of the elaborate structural organization of the cellulosome, which offers an apparent optimal balance amongst modularity, diversity, rigidity and flexibility, is of considerable importance. Thus, characterization of the quaternary structure of the cellulosome has been the focus of many studies. Electron microscopy imaging studies of cellulolytic microbes have revealed dynamic structures on the cell surface that house cellulosomes (7;9). These protuberances form bulbous structures in the absence of substrate that extend and attach to cellulose when it is present (7;9). Small angle X-ray scattering (SAXS) experiments have shown that optimal positioning of enzymes on the substrate requires flexibility in the linker regions separating CohI modules (58;59). In addition, several crystal structures of cellulosomal subunit modules from *C. thermocellum* have been solved including the CBM and the second and seventh CohI modules from CipA, the CohII module from SdbA, and a DocI module (32;34;37;68). More recently, three multimodular structures
have been solved including the CohI:DocI, the X-DocII:CohII, and the CohI$_9$—X-DocII:CohII complexes (25;26;36;Chapter 2).

Here, we describe the structure of a multimodular heterotrimeric cellulosomal complex containing three proteins and five modules, namely the SdbA CohII, the Cel9D DocI, and the CohI$_9$—X-DocII fragment of CipA from *C. thermocellum* studied in solution by SAXS and solved by X-ray crystallography to 2.7 Å resolution, which represents the largest fragment of the cellulosome solved to date. We provide evidence for DocI-dependent inhibition of interscaffoldin contacts and we report the first CohI-DocI structure with a natural binding orientation bias. Finally, our crystallographic and SAXS data indicate that the linker connecting the X-module to the CohI$_9$ is highly flexible; however, surprisingly, flexibility is limited to only 2 dimensional motions.

4.3 Results and Discussion

4.3.1 DocI:CohI$_9$—X-DocII:CohII Crystal Structure

The crystal structure of the DocI:CohI$_9$—X-DocII:CohII complex was solved by molecular replacement using the X-DocII:CohII structure and the CohI$_2$:S45A/S46A DocI mutant structure as search probes (25;36). All molecular replacement strategies that employed wild type CohI:DocI structures failed to provide a solution. The structure was refined to 2.7 Å resolution and the final statistics are summarized in Table 4-1. Four molecules of the complex were found in the asymmetric unit, one of which is shown in (Figure 4-1A). All four structures of the heterotrimeric DocI:CohI$_9$—X-DocII:CohII complex are essentially identical (see below), displaying two well-ordered regions, one
Table 4-1 Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>$P2_12_12_1$</td>
</tr>
<tr>
<td>Cell dimensions $a, b, c$ (Å)</td>
<td>119.37, 186.31, 191.17</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20.0-2.7</td>
</tr>
<tr>
<td>No. unique/free reflections (5%)</td>
<td>107440 (5639)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>95.26 (78.5)</td>
</tr>
<tr>
<td>$R_{sym}$ (%)</td>
<td>4.6 (75.5)</td>
</tr>
<tr>
<td>$I/\sigma I$</td>
<td>30.9 (1.9)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.6 (2.4)</td>
</tr>
<tr>
<td>$R_{work}/R_{free}$ (%)</td>
<td>22.8/27.0</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.243</td>
</tr>
</tbody>
</table>

*R_{sym} = \Sigma |I(k)| - (I)/\Sigma I(k)* where $I(k)$ and $I$ represent the diffraction intensity values of the individual measurements and the corresponding mean values. Values in parentheses are for the highest resolution shell.
Figure 4-1 Crystal structure of the DocI:CohI₉–X DocII:CohII heterotrimeric complex. Panel A is one DocI:CohI₉–X DocII:CohII molecule from the asymmetric unit. The DocI, CohI, X, DocII, and CohII are shown in orange, yellow, pink, green, and blue, respectively. Panel B is an alignment of the X-DocII:CohII region for all four molecules of the DocI:CohI₉–X DocII:CohII complex from the asymmetric unit. Molecules one, two, three, and four of the DocI:CohI₉–X DocII:CohII heterotrimeric complex are coloured green, pink, cyan, and yellow, respectively.
comprising the CipA X-DocII modular pair and the SdbA CohII and the other Cel9D DocI in complex with CipA CohI₉, respectively, connected by a 13-residue linker (Val1687-Lys1699). The complex structures are in extended conformations spanning approximately 150 Å in their longest dimensions.

### 4.3.2 X-DocII:CohII Region

The elongated X-DocII:CohII region, defined by residues 31-195 of the SdbA CohII module and residues 1697-1851 of the X-DocII modular pair from CipA, of the four individual heterotrimeric structures are very similar to one another (backbone RMSD of 0.44 ± 0.05 Å), to the X-DocII:CohII structure (backbone RMSD of 0.60 ± 0.12 Å) (25), and to the analogous region in the CohI₉–X DocII:CohII heterodimeric complex (backbone RMSD of 0.48 ± 0.05 Å) (Chapter 2). The CohII module forms the typical type-II cohesin fold of an elongated nine-stranded β-sandwich in a classical jellyroll topology with a hydrophobic core, a crowning helix, and β-flaps intervening strands 4 and 8 (Figure 4-1A) (25;27;29;30). The DocII binding face is formed by an antiparallel β-sheet including strands 8, 3, 6, and 5 (8-3-6-5 face) (Figure 4-1A). The opposite face is formed by a β-sheet comprising strands 9, 1, 2, 7, and 4 (9-1-2-7-4 face) all in an antiparallel arrangement with the exception of strands 1 and 9, which are parallel (Figure 4-1A). The X-module forms a β-sandwich similar to an Ig-like fold with two antiparallel β-sheets composed of strands 1, 4, and 7, and 2, 3, 5, and 6 with a short α-helix connecting strands 1 and 2 (Figure 4-1A). The DocII is composed of two F-hand loop-helix motifs separated by a 14-residue linker. The two 12-residue loop regions
coordinate one Ca\(^{2+}\) ion each in a pentagonal bipyramid configuration typical of EF-hand Ca\(^{2+}\)-binding proteins.

**4.3.3 DocI:CohI\(_9\) Region**

The CohI\(_9\) forms the typical type-I cohesin fold seen in other type-I cohesin structures (32). The DocI modules are composed of two F-hands, as previously described for the DocII modules, and also structurally coordinate two Ca\(^{2+}\) ions similar to all EF-hand Ca\(^{2+}\)-binding proteins (Fig. 1C). Comparison of the DocI:CohI\(_9\) interaction in the four heterotrimeric structural models show that each DocI module interacts with the expected 8-3-6-5 face of CohI in a single orientation; notably, it is in the same orientation as the Xyn10B S45A/S46A mutant DocI in complex with CipA CohI\(_2\) from *C. thermocellum* (36) and the Cel5A A47S/F48T mutant DocI in complex with the CohI\(_1\) from *Clostridium cellulolyticum* (35) and in a 180° opposite orientation relative to the structure of the wild-type Xyn10B DocI in complex with the CipA CohI\(_2\) module from *C. thermocellum* (26) and the wild-type Cel5A DocI in complex with the CohI\(_1\) from *C. cellulolyticum* (35).

A structural overlay of DocI:CohI\(_9\) region of the four heterotrimeric structures reveals a backbone RMSD of 0.58 ± 0.28 Å, which increases to 1.1 ± 0.28 Å when the previously determined S45A/S46A mutant Xyn10B DocI:CipA CohI\(_2\) complex structure is compared to the four models (26). The most notable difference between the heterotrimeric DocI:CohI\(_9\)–X DocII:CohII structure and the Xyn10B DocI:CipA CohI\(_2\) complex structure involves the linker regions connecting the two helices of the DocI
modules (26). The linker adopts a random coil conformation in the DocI:CohI$_9$–X DocII:CohII heterotrimeric complex structure and is helical in the Xyn10B DocI structure, an observation that may be a result of the helix breaking proline residue in the Cel9D DocI, the larger number of residues in the linker in Cel9D relative to Xyn10B DocI, or may simply arise from differences in crystal packing (26). The Cel9D DocI-CipA CohI$_9$ interface in the heterotrimeric complex is formed by hydrogen bonding contacts and van der Waals contacts (Table 4-2). Analogous residues were previously identified as contributing to the Xyn10B DocI-CipA CohI$_2$ interface (26).

4.3.4 Scaffoldin Dimerization

Inter-scaffoldin contacts involving the X-module of one scaffoldin and a CohI module from another scaffoldin molecule have been reported in the CohI$_9$–X DocII:CohII heterodimer crystal structure (Chapter 2). Furthermore, the CohI$_9$–X DocII:CohII complex is in an equilibrium between the monomeric and dimeric forms in solution, indicating that this phenomenon is more than a result of the local concentration effect of crystallization and may in fact occur in vivo (Chapter 2). With the addition of DocI and under completely different crystallization conditions, the contacts between the X-module of one ternary complex and a CohI, from a symmetry-related complex are conserved (Figure 4-2). The interface involves residues from both the CohI$_9$ and the X-module forming both van der Waals contacts and hydrogen bonds (Table 4-3). The resulting homodimer of the ternary complex displays symmetrical contacts between the X-modules and CohI$_9$ modules of the two ternary complexes, which results in an intertwining of the two scaffoldins, similar to that observed in the CohI$_9$–X DocII:CohII heterodimer (Figure
4-2) (Chapter 2). The intertwining of the CohI–X DocII:CohII heterodimer leaves the DocI binding interface of the CohI available for DocI binding. However, our ternary complex displays more differences than just the bound DocI modules. Comparison of the CohI–X DocII:CohII and DocI:CohI–X DocII:CohII structures reveals a molecular rearrangement of the dimerized complex upon DocI binding, with the CohI–X modular linker acting like a hinge that allows positional flexibility of the X-DocII:CohII region while maintaining the intermolecular CohI–X interface (Figure 4-2).

**4.3.5 Linker Flexibility**

CohI and the X-module of the CipA scaffold are connected by a 13-residue linker (Val1687-Lys1699) (Figure 4-1A). In the CohI–X DocII:CohII structure this linker displays elevated temperature factors, which is indicative of flexibility within this region (Chapter 2). The temperature factors of the linker are also elevated in all four molecules of our DocI:CohI–X DocII:CohII ternary complex structure. Furthermore, alignment of the X DocII:CohII from the four molecules of the DocI:CohI–X DocII:CohII structure reveals slightly different positions of the DocI:CohI (Figure 4-1B). Elevated temperature factors highlight regions of flexibility within our structure. However, the extent of this flexibility may be restrained by crystal packing. For this reason, we have complimented our high resolution crystallographic work with solution scattering studies that will provide insight into the behaviour of ternary complex in solution.
### Table 4-2 Contacts between CipA CohI\textsubscript{9} and Cel9D DocI.

<table>
<thead>
<tr>
<th>Module</th>
<th>Type of Contact</th>
<th>Hydrogen Bonds</th>
<th>van der Waals</th>
</tr>
</thead>
<tbody>
<tr>
<td>CipA CohI\textsubscript{9}</td>
<td></td>
<td>Asp1582, Lys1610,</td>
<td>Ser1580, Asp1613, Ala1615, Met1622,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tyr1617, Glu1629, and</td>
<td>Val1624, Leu1626, Ala1628, and Leu1675</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Glu1677</td>
<td></td>
</tr>
<tr>
<td>Cel9D DocI</td>
<td></td>
<td>Lys546, Lys550, Lys551, Asn570, Ser571,</td>
<td>Lys546, Val549, Leu550, Lys551, Ser571,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ser572, and Arg579</td>
<td>Val574, Thr575, and Ile576</td>
</tr>
</tbody>
</table>
Figure 4-2 Interscaffoldin interactions observed in the DocI:CohI₉–X DocII:CohII heterotrimeric complex. Panel A and panel B display the two homodimers of the DocI:CohI₉–X DocII:CohII heterotrimeric complexes that were observed in the crystal structure. The DocI, CohI, X, DocII, and CohII are shown in orange, yellow, pink, green, and blue, respectively. Symmetry related molecules are shown in grey.
Table 4-3 Interscaffoldin contacts.

<table>
<thead>
<tr>
<th>Module</th>
<th>Type of Contact</th>
<th>Hydrogen Bonds</th>
<th>van der Waals</th>
</tr>
</thead>
<tbody>
<tr>
<td>CipA CohI9</td>
<td></td>
<td>Ile1595, Glu1596, Glu1598, Val1616, Asp1619, and Asp1674</td>
<td>Glu1593, Ile 1595, Glu1596, Ala1615, Val1616, Tyr1717, and Pro1618</td>
</tr>
<tr>
<td>CipA X-module</td>
<td></td>
<td>Glu1696, Gly1697, Tyr1734, Ala1748, Ala1750, and Ser1751</td>
<td>Asn1698, Glu1745, Tyr1747, Tyr1781, Val1783, Ala1797, and Ala1799</td>
</tr>
</tbody>
</table>
4.3.6 SAXS Studies

SAXS was used to investigate the structure of the DocI:CohI$_9$–X DocII:CohII complex in solution, including the flexibility of the CohI$_9$–X-module linker. The SAXS profile (Figure 4-3A) and linear radius of gyration ($R_G$) of the DocI:CohI$_9$–X DocII:CohII complex indicate that the complex is well behaved and aggregation-free in solution over a wide range of concentrations (2-14 mg/mL). The maximal dimension ($D_{\text{max}}$) of the ternary complex in solution is 150 Å, which is consistent with the extended length of the DocI:CohI$_9$–X DocII:CohII crystal structure. An ensemble of twenty calculated SAXS envelopes displayed only slight topological variations and was used to generate an average envelope, which reveals a large and a small domain connected by a short neck (Figure 4-3B). Attempts at manual superpositioning the X-DocII:CohII and the CohI:DocI crystal structures onto the two domains, suggest that the large and small domains comprise the X-DocII:CohII complex and the CohI:DocI complex, respectively (Figure 4-3B). Consistent with the crystal structure, the two globular domains of the SAXS envelope are separated by an extended linker.

In order to investigate the degree of flexibility of the CohI$_9$–X module linker in solution we utilized the BILBOMD rigid body modelling strategy, which employs molecular dynamics (MD) to generate thousands of different conformers, from which SAXS curves can be calculated and compared against experimental data (84). The best-fit model from the pool of calculated conformers shows an extended conformation consistent with the crystal structure as well as the ab initio SAXS envelope (Figure 4-3B). Flexible protein systems are not represented well by a single model. For this reason,
Figure 4-3 SAXS analysis of the DocI:CohI₉–X DocII:CohII heterotrimeric complex. Panel A displays a plot of the raw SAXS data. Panel B contains the ab initio model that was generated. The CohI₉:DocI and X-DocII:CohII fragments from the crystal structure were manually placed within the SAXS envelope. Panel C contains the four different conformers selected by the BILBOMD analysis aligned over the X-DocII:CohII region. The DocI, CohI, X, DocII, and CohII are shown in orange, yellow, pink, green, and blue, respectively.
we used a genetic algorithm-based minimal ensemble search (MES) to identify the minimum number of conformers required to optimally fit our data (84). By this means we select conformers that likely exist in the population and will better represent the flexibility of the system than a single model. By assembling an ensemble of four selected conformers the fit to the experimental data was improved (Figure 4-3C). Interestingly, the minimal ensemble of conformers displays domain motion in only two dimensions. Although more conformations are present in the population, the minimal ensemble displays those conformers that best represent the population. As such, this may indicate restrained motion of the linker instead of random movement, which may position enzymes in order to establish or maximize enzymatic synergy. Linker flexibility has been shown to be important in optimal positioning of enzymes on the substrate (58;59). The apparent flexibility between CohI_9 and the X-module may play a similar role. Our SAXS experimental results (Figure 4-3C) provide direct evidence of the flexible linker, which would otherwise not be possible from the crystallographic structural studies.

Our solution scattering studies indicate that the DocI:CohI_9–X DocII:CohII ternary complex is monomeric in solution under the buffering conditions and at concentrations at which homodimers of the CohI_9–X DocII:CohII complex exist (Chapter 2). However, since CohI_9–X module interface contacts still exist at much higher concentrations in the crystal, DocI binding does not prevent but may rather weaken the interscaffoldin interaction. The nature of DocI-dependent inter-scaffoldin binding inhibition is unclear since the DocI modules appear to interact only with the 8-3-6-5 face of the CohI_9 modules in the crystal structure. However, CohI_9–X module mediated inter-
scaffoldin interactions could play a role in cellulosome assembly whereby the intertwined scaffoldin structure ensures access to the 8-3-6-5 face of the enzyme-free CohI₉ modules. Upon enzyme binding, the interaction between scaffoldins is weakened to allow optimal positioning of enzymes on the substrate. The inter-scaffoldin interactions observed in the CohI₉–X DocII:CohII heterodimer and the DocI:CohI₉–X DocII:CohII heterotrimer crystal structures may indicate an important role for scaffoldin-scaffoldin interactions in cellulosome function. Indeed, in nature, the cellulosomes of this bacterium are housed in a highly concentrated state in cell-surface protuberance-like structures, and their native microenvironment likely emulates a condition somewhere between with the crystalline and the solution states observed here. Interscaffoldin interactions may increase synergy through the enzyme proximity effect or may mimic the branching seen in cellulosomes that have several scaffoldin subunits.

Our data suggest the first region of flexibility of the cellulosomal scaffoldin extending from the cell surface is located in the linker between the CohI₉ and X module. These results indicate restrained and highly selective flexibility is optimal for cellulosome function.

4.4 Materials and Methods

Cloning, protein expression, purification, and crystallization of the SdbA CohII, Cel9D DocI, and CipA CohI₉—X-DocII complex along with diffraction data collection, indexing were performed as previously described (Chapter 3). The structure was solved by molecular replacement using PHASER (71). Manual fitting of the model was carried
out using COOT and XFIT and refined with REFMAC (73). RMSD values were calculated using DALILITE (85).

Protein samples used for SAXS experiments were prepared as previously described (Chapter 2). Lysozyme and BSA standards were used to calibrate the I/(0) values and assess potential aggregation of samples. Protein concentrations were determined using $A_{280}$ from relative molecular mass and Abs0.1% (=1g/L) parameters calculated using ProtParam (86). $A_{280}$ measurements were made using a NanoDrop ND-1000 spectrometer (NanoDrop Technologies). SAXS data were measured using an Anton Paar SAXSess line-collimation instrument with a sealed tube source and 1D elliptical focusing graded multilayer mirror optics (d spacing = 4 ± 0.07 nm) with a Max-Flux Osmic block collimator (flatness, < 14 µm/cm). The scattering intensities from each protein sample and its final dialysate were recorded for 120 min at 4 ºC using 10-mm slit beam geometry. Data were recorded using 2D position-sensitive image plate (42.3 × 42.3 µm² pixel size; sample-to-image plate distance, 264.5 mm). Image plate was read using a Perkin Elmer Cyclone phosphor storage system driven with OptiQuant software. Anton Parr SAXSquant2D software was used to average the 2D images over a 7-mm integration width to produce the slit-smeared 1D-scattering profiles as I(q) versus q for $q \approx 0.007–0.6$ Å$^{-1}$ [$q = ((4\pi/\lambda) \sin \theta$, $\lambda = 1.5418$ Å (Cu Kα), $\theta$ = half the scattering angle]. The program SAXSquant1D (Anton Paar, Austria) was used to subtract the scattering of each solvent blank from that of each protein solution to yield the scattering profiles of the protein molecules alone. Low resolution $ab$ $initio$ shapes were generated with GNOM output files and DAMMIN (87;88). Twenty models were averaged using the DAMAVER
software package and presented as a filtered model (89). Rigid body modelling was performed using BILBOMD (84). Molecular dynamics (MD) was used to explore the conformational space adopted by the DocI:CohI—X DocII:CohII complex. A Minimal Ensemble Search (MES) was used to identify the minimal ensemble of conformations required to best fit the experimental data (84). Structural figures were prepared using PyMOL.
Chapter 5

Solution Structure and Dynamics of the Cellulosome from *Clostridium thermocellum* Revealed by SAXS

5.1 Abstract

*Clostridium thermocellum* produces the prototypical cellulosome, a multienzyme cell surface bound complex with efficient cellulose adhering and degrading properties. The cellulosome is composed of a non-catalytic scaffoldin, enzymatic subunits, and an anchoring protein. Each protein component contains multiple globular modules separated by linkers, which are thought to impart flexibility to the system. However, the precise spatial arrangement of various components, flexibility of the linkers, and their impact on the modular arrangement of the cellulosome are poorly understood. Here, we describe the structures and solution dynamics of representative enzyme-bound and enzyme-free sections of the cellulosome scaffoldin CipA from *C. thermocellum* using small angle X-ray scattering (SAXS) together with available high resolution structural information of cellulosome modules. We show that scaffoldin linker flexibility is highly restricted, which may suggest interactions between the linkers and modules or between neighbouring modules, or potentially previously unidentified structure within the linkers themselves. Moreover, this study provides insight into enzyme arrangement on the cellulosomal scaffoldin. Neighbouring enzyme subunits are rotated by $\sim$90° relative to the previous associated enzyme on the scaffoldin. Catalytic subunits maintain close proximity to their cognate dockerin module, which may be caused again by contacts between the linker and either the enzymatic module or dockerin module, direct
interaction of the dockerin and enzymatic module, or structure within the linker region itself. This study provides an unprecedented view of modular arrangement and flexibility within the cellulosome and transforms the traditional view of a highly flexible cellulosome into a complex of restrained and in some cases very limited flexibility.

5.2 Introduction

Cellulose, the major constituent of plant cell walls, is the most abundant organic molecule on Earth. With increasing energy consumption, the depletion of fossil fuels, and growing environmental concerns, development of alternative fuel sources is paramount (90). Conversion of plant biomass to ethanol represents an inexpensive, renewable, and environmentally friendly alternative to fossil fuels (90). The critical bottleneck that is preventing bioethanol from becoming competitive as an energy source is the process of hydrolysing the plant cell wall into its individual sugar components (91). Thousands of microbial species have been discovered that can rapidly and efficiently degrade plant cell wall polysaccharides through the synergistic activity of a variety of cellulases, hemicellulases, and other hydrolases all housed in large cell surface bound protein complexes called cellulosomes (92).

The best studied cellulosome from the thermophilic anaerobe Clostridium thermocellum is composed of three main components: a multimodular non-catalytic scaffoldin subunit called CipA, a variety of DocI containing catalytic subunits including cellulases, hemicellulases, and other hydrolytic enzymes, and one of three type-II cohesin module (CohII) containing cell surface bound proteins SdbA, Orf2p, or OlpB. The CipA
scaffoldin is comprised of a cellulose-specific binding module (CBM), an X-module (X) of unknown function, a type-II dockerin module (DocII), and nine type-I cohesin modules (CohI) all connected by flexible linkers of varying lengths (13;14). The enzymatic subunits bind tightly to the CohI modules of CipA via type-I dockerin modules (DocI) (22;26;36). The CipA scaffoldin is anchored to the cell surface bound proteins through its DocII module (19;57).

Detailed structural information of cellulosomes will provide rationale for designer cellulosome engineering and enhanced bioethanol production (93). Several representative high resolution structures of individual modules have been solved using X-ray crystallography and NMR, including the CBM (68), the second and seventh CohI modules (32;34), the DocI module (37), and the SdbA CohII module (27). In addition, several high resolution multimodular structures have been solved including the CohI:DocI (26;35;36), and the X-DocII:CohII (25), CohI₉—X-DocII:CohII (Chapter 2), and the DocI:CohI₉—X-DocII:CohII (Chapter 4). SAXS studies performed previously have shown that flexibility in engineered cellulosomes is required for optimal enzyme positioning (59). Electron microscopy images have revealed the dynamic nature of the cellulosomal ultrastructure (7;9). In the absence of substrate the cellulosome forms bulbous protuberances on the surface of the cell (9). However, in the presence of a crystalline cellulose substrate, the cellulosome forms fibrous arm-like structures that adhere to the surface of the cellulose (9).

In the current study, SAXS was used to investigate the structure and flexibility of several representative fragments of the cellulosome from *C. thermocellum* with and
without enzymes bound. The flexibility of the fragments was found to be highly restricted. Notably, the scaffoldin linkers exhibit motion that is limited to a single plane. Moreover, DocI and the enzymatic subunit of the Cel8A enzymes maintain close contact in all of the enzyme-bound structures and adjacent enzymes on the scaffoldin are positioned about 90° from neighbouring enzymes. This study suggests that selective flexibility is an inherent part of and critical for proper cellulosome function.

5.3 Materials and Methods

5.3.1 Protein Preparation

CohI₈—CohI₉—X, CohI₈—CohI₉—X-DocII, and CohII were all expressed and purified and complexes were formed as previously described for CohII (69), and the CohI₉—X-DocII:CohII and DocI:CohI₉—X-DocII:CohII complexes (Chapter 4). However, CohI₈—CohI₉—X and CohI₈—CohI₉—X-DocII were expressed from a pET21 vector. pET28a expression vectors for the S₄₅₈A/S₄₅₉A Cel8A enzyme alone and co-expressed with CohI₃—CohI₄—CohI₅ or CohI₂—CBM—CohI₃ were provided by Dr. Carlos Fontes from Universidade Técnica de Lisboa, Portugal. These constructs were transformed into Tuner cells (Invitrogen). A single colony was used to inoculate a 25 mL Luria Bertani (LB) broth overnight culture supplemented with 100 µg/mL kanamycin. The overnight culture was grown overnight at 37°C and used to start a 1L LB culture supplemented with 100 µg/mL kanamycin. Cultures were grown to an OD₆₀₀ of 0.6, induced with 0.2 mM IPTG, and grown overnight at 19°C. Cells were harvested by centrifugation for 20 minutes at 3000 g and then resuspended in Buffer A (20 mM Tris,
pH 8.0, 300 mM NaCl, 10 mM imidazole, and 1 mM CaCl$_2$). Lysozyme (1 mg/mL) was added to the suspension and incubated on ice with stirring for 30 minutes. The cells were further lysed by sonication and the lysate was cleared by centrifugation at 16 000 g using a JA20 rotor. Ni$^{2+}$ affinity resin (2 mL) was added to the supernatant and incubated at 4° with rocking for 1 hour. The mixture was subsequently applied to a column and washed with Buffer A and eluted with Buffer A with 400 mM imidazole added. The elution fractions were concentrated and applied to a Sephadex 200 column (Amersham) that was pre-equilibrated with Buffer B (50 mM HEPES, pH 7.5, 50 mM NaCl, 1 mM CaCl$_2$, and 1 mM DTT). Fractions containing protein were analyzed by SDS-PAGE and fractions containing pure complexes were pooled and concentrated. The S458A/S459A Cel8A enzyme-bound CohI$_8$—CohI$_9$—X-DocII:CohII complex was formed by mixing purified CohI$_8$—CohI$_9$—X-DocII:CohII complex and Cel8A followed by size exclusion chromatography in Buffer B. Protein containing fractions were analyzed by SDS-PAGE and fractions containing pure complex were pooled and concentrated (Figures 5-1 to 5-5).

5.3.2 SAXS Data Collection and Analysis

SAXS data was collected and used to generate ab initio structures and rigid body structures as described in Chapter 4. The Ensemble Optimization Method (EOM) was performed as described previously (94). Molecular weight calculations and Guinier plot analysis of SAXS data confirmed that each sample was monomeric and aggregation-free (Figure 5-1 to 5-5). Structural parameters calculated from SAXS data are tabulated in Table 5-1.
Figure 5-1 CohI₈—CohI₉-X protein purification and SAXS data. A. SDS-PAGE of purified protein. B. Raw SAXS data. C. Guinier plot.
A

KDa
75 50 37 25 20 15

CohI\textsubscript{8}-CohI\textsubscript{9}-X-DocII
CohII
Figure 5-2 CohI$_8$—CohI$_9$-DocII:CohII protein purification and SAXS data.  A. SDS-PAGE of purified complex.  B. Raw SAXS data.  C. Guinier plot.
A

KDa

75
50
37
25
20
15
10

Cohl₈-Cohl₉-X-DocII
Cel8A

CohII
Figure 5-3 Cel8A:CohI₅—CohI₇-X-DocII:CohII protein purification and SAXS data. A. SDS-PAGE of purified complex. B. Raw SAXS data. C. Guinier plot.
Figure 5-4 Cel8A:CohI₃—CohI₄—CohI₅ protein purification and SAXS data. A.
SDS-PAGE of purified complex. B. Raw SAXS data. C. Guinier plot.
A

<table>
<thead>
<tr>
<th>KD</th>
<th>75</th>
<th>50</th>
<th>37</th>
<th>25</th>
<th>20</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cohl&lt;sub&gt;2&lt;/sub&gt;-CBM-Cohl&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Cel8A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5-5 Cel8A:CohI₂—CBM—CohI₃ protein purification and SAXS data. A. SDS-PAGE of purified complex. B. Raw SAXS data. C. Guinier plot.
**Table 5-1 Structural parameters from SAXS data.**

<table>
<thead>
<tr>
<th>Complex</th>
<th>RG (Å)</th>
<th>Dmax (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CohI₈—CohI₉—X</td>
<td>39.70</td>
<td>120.0</td>
</tr>
<tr>
<td>CohI₈—CohI₉—X-DociI:CoII</td>
<td>52.90</td>
<td>180.0</td>
</tr>
<tr>
<td>Cel8A:CohI₈—CohI₉—X-DociI:CoII</td>
<td>67.60</td>
<td>205.6</td>
</tr>
<tr>
<td>Cel8A:CohI₃—CohI₄—CohI₅</td>
<td>70.30</td>
<td>207.0</td>
</tr>
<tr>
<td>Cel8A:CohI₂—CBM—CohI₃</td>
<td>66.40</td>
<td>210.0</td>
</tr>
</tbody>
</table>
5.3.3 Molecular Modelling

Molecular models were generated using the Phyre server and PyMOL (95).

5.4 Results and Discussion

5.4.1 Ab initio Models of Enzyme-Free Cellulosomal Fragments

In order to gain initial structural insights into the cellulosomal fragments, first we used the \textit{ab initio} methods. Specifically, structures of two enzyme-free constructs, CohI\textsubscript{8}—CohI\textsubscript{9}—X and CohI\textsubscript{8}—CohI\textsubscript{9}—X-DocII:CohII, were generated using DAMMIN (88). The envelopes of both CohI\textsubscript{8}—CohI\textsubscript{9}—X and CohI\textsubscript{8}—CohI\textsubscript{9}—X-DocII:CohII constructs (Figure 5-6 and Figure 5-7) look very similar. However, the larger construct, CohI\textsubscript{8}—CohI\textsubscript{9}—X-DocII:CohII, is slightly longer, which can be accounted for by the additional type-II interaction modules and used to help orient the structure. These structures both contain three separate regions that correspond to the three rigid domains found in the constructs, CohI\textsubscript{8}, CohI\textsubscript{9}, and either the X-DocII:CohII or the X-module alone. Each of these domains is oriented at an angle of roughly 90° to one another with the two terminal domains pointed 180° from one another in both structures. This suggests that no interactions occur between the terminal domains of these two constructs. There is potential for interactions between the central CohI\textsubscript{9} module and the two terminal domains that help stabilize the modular orientations. Although these structures are not uniformly thick across their entire length, there are no clear boundaries between domains and linkers, which may appear to suggest some flexibility. However, manual super-positioning of the three domains within these structures suggests that the
Figure 5-6 SAXS based ab initio structure of CohI$_8$—CohI$_9$—X. Panel A displays two views, rotated 90° from one another, of the ab initio structure of CohI$_8$—CohI$_9$—X. Panel B shows the same two views of the CohI$_8$—CohI$_9$—X ab initio structure shown in Panel A with the crystal structure of CohI$_9$ and the X-module as well as a model of the CohI$_8$ manually placed within the envelope (25;Chapter 2). The CohIs and the X-module are shown in yellow and pink, respectively. The envelope is shown in grey.
Figure 5-7 SAXS based ab initio structure of CohI$_8$—CohI$_9$—X-DocII:CohII. Panel A displays two views, rotated 90° from one another, of the ab initio structure of CohI$_8$—CohI$_9$—X-DocII:CohII. Panel B shows the same two views of the CohI$_8$—CohI$_9$—X-DocII:CohII ab initio structure shown in Panel A with the crystal structure of CohI$_9$ and the X-DocII:CohII as well as a model of the CohI$_8$ manually placed within the envelope (25;Chapter 2). The CohIs, X-module, DocII, and CohII are shown in yellow, pink, green, and blue, respectively. The envelope is shown in grey.
linkers may be present in a compact arrangement since the three domains alone take up almost the entire length of the SAXS envelopes.

5.4.2 Ab initio Models of Enzyme-Bound Cellulosomal Fragments

Similar to the enzyme-free constructs described above, ab initio structures of three representative multimodular portions of the CipA scaffoldin subunit including CohI2—CBM—CohI3, CohI8—CohI9—X-DocII:CohII, CohI3—CohI4—CohI5 and, all with Cel8A enzymes bound, were also determined (Figure 5-8, Figure 5-9, and Figure 5-10). The Cel8A sequence was mutated in order to incorporate S458A/S459A mutations that ensure DocI binding into a single orientation to help maximize sample homogeneity. Both the enzyme associated CohI8—CohI9—X-DocII:CohII and the CohI3—CohI4—CohI5 constructs appear as bent (Figure 5-9 and Figure 5-10) structures lacking two globular domains. The CohI8—CohI9—X-DocII:CohII structure contains three globular domains, whereas the CohI3—CohI4—CohI5 structure is composed of four separate globular domains. The missing domains may be too flexible to appear in the ab initio structures. In the absence of additional overlapping fragments it is difficult to determine which domain corresponds to which modules in each of these structures. The CohI2—CBM—CohI3 structure is composed of three globular domains. The two terminal domains are oriented about 180° from one another, which may indicate a role for the CBM module in enzyme positioning.
Figure 5-8 SAXS based *ab initio* structure of CohI$_2$—CBM3—CohI$_3$ bound to S458A/S459A mutant Cel8A enzymes. Panel A displays two views, rotated 90° from one another, of the *ab initio* structure of the CohI$_2$—CBM3—CohI$_3$ S458A/S459A Cel8A complex. Panel B shows the same two views of the *ab initio* structure shown in Panel A with the crystal structure of CohI$_2$ and models of Cel8A enzymes and CohIs manually placed within the envelope (26;36). The CohIs and the enzymes are shown in yellow and orange, respectively. The envelope is shown in grey.
Figure 5-9 SAXS based *ab initio* structure of CohI₈—CohI₉—X-DocII:CohII bound to S458A/S459A mutant Cel8A enzymes. Panel A displays two views, rotated 90° from one another, of the *ab initio* structure of the CohI₈—CohI₉—X-DocII:CohII S458A/S459A Cel8A complex. Panel B shows the same two views of the *ab initio* structure shown in Panel A with the crystal structure of the X-DocII:CohII and CohI₉ and models of Cel8A enzymes and the CohI₈ manually placed within the envelope (25, Chapter 2). The CohIs, CohII, DocII, X-module, and the enzymes are shown in yellow, blue, green, pink, and orange, respectively. The envelope is shown in grey.
Figure 5-10 SAXS based *ab initio* structure of CohI$_3$—CohI$_4$—CohI$_5$ bound to S458A/S459A mutant Cel8A enzymes. Panel A displays two views, rotated 90° from one another, of the *ab initio* structure of the CohI$_3$—CohI$_4$—CohI$_5$ S458A/S459A Cel8A complex. Panel B shows the same two views of the *ab initio* structure shown in Panel A with the models of Cel8A enzymes and CohIs manually placed within the envelope. The CohIs and the enzymes are shown in yellow and orange, respectively. The envelope is shown in grey.
5.4.3 Flexibility of an Enzyme-Free Scaffoldin Fragment by the Ensemble Optimization Method

Although the SAXS-based single model method has been widely used and is the most popular SAXS approach, *ab initio* structures may not represent flexible proteins well since the result is a single structure and a flexible protein is capable of assuming many different conformations (94). Moreover, Bernadó has shown that *ab initio* SAXS structures of flexible proteins result in uniformly thick cigar or worm shaped structures, which are difficult to interpret (94). Hence, the limitations of traditional *ab initio* methods have lead to the development of two new tools for analyzing flexible proteins with SAXS, the Ensemble Optimization Method (EOM) and BILBOMD (84;94). EOM uses a high resolution structure determined by NMR, X-ray crystallography, or modelling to calculate ~10 000 conformers based on user defined segments of rigidity and flexibility. A scattering curve is generated for each conformer and then a genetic algorithm is used to determine the optimal mixture of these different conformers to fit the experimental data. The data is binned based on $R_G$ and $D_{max}$ for both the original 10 000 conformers and the experimentally selected ensemble for comparison. The EOM method was applied to the CohI$_8$—CohI$_9$—X segment of the CipA scaffoldin. The conformational sampling indicates that the unrestricted molecule would have a $D_{max}$ of about 175 Å whereas the ensemble has a $D_{max}$ of about 150 Å. This indicates that the CohI$_8$—CohI$_9$—X molecule is more compact than the conformational pool, which may be caused by inter-modal interactions, linker-module interactions, or potentially previously unidentified structure or structural restrictions within the linkers themselves.
5.4.4 Flexibility of Enzyme-Free Scaffoldin Fragments by BILBOMD

One limitation of EOM is that it cannot be applied to protein-protein complexes. Therefore, the remainder of the flexibility analysis for this study was performed using BILBOMD. BILBOMD also begins with conformational sampling, which yields ~10 000 conformers. However, in the case of BILBOMD, the conformational sampling is performed by Molecular Dynamics (MD). This allows analysis of protein complexes, since in addition to regions of rigidity and flexibility, protein-protein contact regions can also be defined. Scattering curves are generated for the pool of conformers and selected for fit to the experimental data based on a genetic algorithm. However, BILBOMD will determine the single best fit conformer as well as a minimal ensemble of five or less conformers that when combined best fit the experimental scattering curve. BILBOMD was applied to enzyme-free CohI₈—CohI₉—X section of the CipA scaffoldin (Figure 5-11). A minimal ensemble of four conformers was generated that best fit the SAXS data. Interestingly, the linker region between the CohI₀ and the X-module displays the same two dimensional motion described for the DocI:CohI₀—X-DocII:CohII structure (Chapter 4). However, the range of motion is larger. The linker between the two CohIs displays less order between each of the conformers in the minimal ensemble.

5.4.5 Flexibility of Enzyme-bound Scaffoldin Fragments by BILBOMD

BILBOMD analysis was also performed on the CohI₈—CohI₀—X-DocII:CohII fragment of CipA bound to S458A/S459A Cel8A enzymes (Figure 5-12). In this case, the minimal ensemble is composed of three conformers. The same two dimensional motions observed between the CohI₀—X linker exist. However, the range of motion is less than
Figure 5-11 SAXS selected minimal ensemble of Coh1₈—Coh1₀–X. Panel A and Panel B displays the alignment of the four conformers selected by BILBOMD for the Coh1₈—Coh1₀–X aligned over the Coh1₀ module in two different views, rotated 90° from one another (84). The Coh1₈s and the X-module are shown in yellow and pink, respectively.
Figure 5-12 SAXS selected minimal ensemble of the CohI₈—CohI₉—X-DocII:CohII S458A/S459A Cel8A complex. Panel A and Panel B displays the alignment of the three conformers selected by BILBOMD for the CohI₈—CohI₉—X-DocII:CohII bound to S458A/S459A Cel8A enzymes aligned over the X-DocII:CohII region in two different views, rotated 90° from one another (84). The CohIs, X-module, DocII, CohII, and the enzymes are shown in yellow, pink, green, blue, and orange, respectively.
that seen in the DocI:CohI$_9$—X-DocII:CohII structure (Chapter 4). It is not clear what role this restricted motion has in cellulosome function. However, this may help guide the enzymatic subunits to their substrate. The position of the GH8—DocI:CohI$_8$ segment relative to the GH8—DocI:CohI$_9$ varies only by how extended the linker is between the two type I cohesins. Continuous extending and contracting of the scaffoldin may also be involved in locating substrate or next cleavable bond in the sugar polymer. The angle between the two enzymes is about 90° in all three conformers. This angle may help to maximally explore the extracellular space for cellulose or optimally position the enzymes in the complex matrix of the plant cell wall. Interestingly, the enzymatic subunits are in contact with the DocI modules for all of the enzymes in all three conformers of the ensemble despite defining the 42-residue connecting linker as flexible. Alignment of the enzymatic subunits from the three conformers reveals only small differences in the relative position of the DocI modules. This indicates that the majority of cellulosomal flexibility occurs within the scaffoldin subunit inter-cohesin linkers.

Structural biology provides unique mechanistic insight about complex protein functions. Large size and inherent flexibility of cellulosomes and other complexes can hinder traditional structural characterization and, as a result, limit our understanding of molecular function. The current study provides rationale to examine large flexible complexes by combining high resolution data with SAXS analysis. Investigation of additional segments of the cellulosome by this approach will help to build a working
model of cellulosome flexibility, motions, and three dimensional modular arrangements, which can ultimately be used to help understand cellulosome function.
Chapter 6
General Discussion

6.1 Forward

Most investigations of cellulosomal structure to date have focused on obtaining high resolution structural information of individual modules or Coh:Doc modular pairs. Although these studies have been instrumental in shaping the current view of the cellulosome function and assembly, they have not been able to explain the enzymatic synergy of the cellulosome. As a result, we have a detailed view of how individual components of the cellulosome function in isolation, but not how they function together. The studies described in this thesis shifts the focus to structural studies and dynamics of larger multi-modular sections of the cellulosome. This chapter will discuss the contributions these studies have made as well as future directions for cellulosome research.

6.2 General Findings and Their Implications

The Coh$_9$—X-DocII:CohII and the DocI:Coh$_9$—X-DocII:CohII complex structures represent the two largest high resolution structures of the cellulosome solved by X-ray crystallography to date. Both exhibit the same novel inter-scaffoldin interaction between the Coh$_9$ module of one CipA molecule and the X-module of adjacent CipA molecule within the crystals, which appears only in the absence of DocI in solution. The local concentration of cellulosomal scaffoldins may be sufficiently high in cell surface polycellulosome containing protuberances to support inter-scaffoldin interactions.
Therefore, the inter-scaffoldin interactions that are observed in the two crystal structures presented here may play a potentially critical role in cellulosome assembly or higher order organization.

The DocI:CohI_9—X-DocII:CohII complex structure provides the first evidence for a DocI orientation bias. All four DocI modules are bound to their respective CohI binding partner in an orientation opposite to other wild-type CohI:DocI structures. The preference for this orientation could arise from small but significant differences between the CohI_9 and Cel9D DocI compared to other CohI:DocI structures that have been reported. However, it is also possible that the chemical composition of the crystallization condition itself favours this orientation.

Using SAXS an envelope was calculated for the DocI:CohI_9—X-DocII:CohII, CohI_8—CohI_9—X-DocII:CohII, and the CohI_8—CohI_9—X scaffoldin fragment without enzymes bound. In addition, SAXS envelopes were calculated for the CohI_8—CohI_9—X-DocII:CohII, CohI_3—CohI_4—CohI_5 and CohI_2—CBM—CohI_3 fragments with enzymes bound. Each of these provides a general shape for the molecules and relative positioning of various modules.

SAXS was also used to investigate the flexibility of the cellulosomal fragments. This analysis revealed constraints in the flexibility of several linkers within the scaffoldin. It was noted that the Cel8A enzymatic subunits remain in very close proximity to their cognate type-I dockerins, despite long linkers connecting the two. In addition, enzyme positioning on the scaffoldin rotates approximately 90° relatively to
neighbouring enzymes. All of these structural restraints were discovered in what were previously thought to be unrestrained and highly flexible linker regions within the cellulosome. Further studies will be required in order to determine the reason for these structural restraints.

6.3 Inter-scaffoldin Interactions

EM studies have revealed that cellulosomes exist in clusters on the surface of the cell in either bulbous or protracted protuberances (7;9). Also, two anchoring proteins have been identified in *C. thermocellum* that have more than one type-II cohesin linked by as few as nine residues. Both of these observations suggest that cellulosome and scaffoldin proximity may be important for proper cellulosome function. The two structures of the two proteins presented in this thesis, the CohI<sub>9</sub>—X-DocII:CohII and the DocI:CohI<sub>9</sub>—X-DocII:CohII, were crystallized under different conditions but both reveal identical contacts between the X-module and a CohI<sub>9</sub> module from another scaffoldin molecule (Chapter 2;Chapter 3). Sedimentation velocity experiments confirm the presence of CohI<sub>9</sub>—X-DocII:CohII dimers in solution as well albeit only a small percentage of the overall population. The effect of proximity has been established in cellulosomal enzymes using engineered cellulosomal scaffolds: two enzymes in solution display lower cellulolytic activity relative to those same two enzymes at the same concentration bound to an engineered scaffoldin. Therefore, the observed inter-scaffoldin interactions may simply be another means to increase local concentration of cellulolytic enzymes. More complex cellulosomes have been discovered in other organisms such as *Acetivibrio cellulolyticus, Bacteroides cellulosolvens,* and *Ruminococcus flavefaciens*.
that employ multiple scaffoldin proteins consisting of a primary scaffoldin that interacts with many enzyme binding secondary scaffoldin subunits (24;96;97). This multiple scaffoldin branching structure is not present in \textit{C. thermocellum}. However, its prevalence in nature suggests that it offers some benefit. The inter-scaffoldin interactions described here may mimic this structure in the cellulosome of \textit{C. thermocellum}, thereby providing a similar enzymatic advantage. Furthermore, the X-module is present in several cellulosome and non-cellulosome systems in a variety of different positions and numbers. However, the function is still unclear. The X-module makes significant contacts with type-I cohesins from neighbouring scaffoldins in the two crystal structures presented here, which suggests the X-module may be a protein-protein interaction module. X-module-CohI interactions may be critical for the higher order assembly of cellulosome and non-cellulosomal quaternary structures. Despite these intriguing possibilities, further study is needed to confirm the presence of inter-scaffoldin interactions in cellulosomes as well as confirm the possible involvement of the X-module in these interactions.

\subsection*{6.4 Type-I Cohesin-Dockerin Dual Binding Preferences}

The signature duplicated sequence of many DocI modules creates a symmetrical dyad structure creating two equivalent binding surfaces. Symmetric DocI modules therefore have two binding modes that may allow for plasticity during enzyme assembly on to their repetitive cognate CohI scaffolds. Carvalho \textit{et al.} found that mutagenesis of one binding surface causes the DocI to preferentially bind its CohI with second equivalent surface and vice versa (36). Using this knowledge the structures and energetics of both interfaces have been previously studied in \textit{C. thermocellum} and \textit{C.}
cellulolyticum (35;36). In *C. thermocellum* the Xyn10B DocI binding modes are related by a rotation of 180° creating two interfaces of nearly identical contact sets that result in the same ∆G but arise from different enthalpy and entropy contributions. In *C. cellulolyticum*, the DocI binding surfaces were also related by rotation of 180°, which created two similar interfaces but resulted in different ∆G (22). Mutant A16S/L17T resulted in approximately half the $K_A$ compared to the equivalent mutant of A47S/F48T (22). This data suggests that the A47S/F48T is the favored orientation for Cel5A DocI under those experimental conditions. A similar observation was observed for the CelD DocI in *C. thermocellum*, where the SXXM/TXXQ mutation resulted in approximately half the $K_A$ compared to the equivalent SXXL/SXXL mutation (22). To date, the structural models have poorly explained the clear differences in the enthalpy and entropy contributions. In the asymmetric unit of the trimeric complex all four CelD DocI molecules use the same docking mode to bind the ninth CohI of CipA. The DocI makes no crystal contacts in this lattice removing any possibility of induced orientation bias. To rationalize the consistency of the CelD orientation we evaluated the symmetry of the DocI. By rotating overlaying the CelD DocI on both orientations of the Xyn10B DocI, we found that unlike Xyn10B DocI the C-terminal of loop of helix 3 and the C-terminal loop of helix 1 form unique structural configurations. This asymmetry creates a distinct set of intermolecular contacts that could easily lead to the differences in $K_A$, $\Delta H$ and $T\Delta S$ measured for two orientations (22). If the energetics of one orientation are favoured in the isotonic solutions of the ITC experiment, then it may be possible that there exists chemical conditions for which the second orientation is favoured. In our structure the
DocI modules all use an orientation that is less favorable in isotonic solution but has a larger entropic contribution. The bias of this orientation is likely driven by the high salinity of the LiSO₄ precipitant. The unmutated Xyn10B DocI in isotonic crystal conditions favour the opposite orientation. This is the first evidence for chemical bias of the DocI orientation.

We propose a new function of the DocI dual binding mode as a compensating mechanism for environmental fluctuations. The Coh:Doc interactions have coevolved to create high affinity associations in an environment with unlimited inconsistencies of temperature and salinity. It has been shown that DocI shows a near perfect enthalpy/entropy compensation over a large range of temperatures. The duality of the DocI binding mode may allow for a CohI:DocI to adapt to these changes by sustaining two docking modes with differing contributions of enthalpy and entropy.

6.5 Intermodular Linkers and the Cellulosome

All components of cellulosomes are modular proteins with linkers of varying lengths separating each module. Most studies of cellulosome structure and function focus on the modular regions of these proteins. However, studies described here and others illustrate the importance of the linker regions in cellulosome structure and function (58;59;Chapter 4;Chapter 5). We show that linker regions of recombinant cellulosome segments display restrained flexibility. Recent studies by other groups suggest that both linker length and flexibility impact cellulolytic activity in designer cellulosomes (43;59). Several genome sequences are available from cellulosome producing bacteria, which
have revealed the diversity within these regions of different cellulosome components and complexes. Inter-modular linkers range in size from five to 721 residues. The longer linkers tend to be found on anchoring proteins and shorter linkers are typically found on the scaffoldin subunits. In some cases, longer linkers contain internal repeats that may suggest structuring within these regions. Many scaffoldins and anchoring protein linkers are rich in proline and threonine residues; however, others, such as the *C. cellulosorans* CipC scaffoldin, show no preference. In addition to length and amino acid composition, cellulosomal linkers are also highly glycosylated on threonine and possibly serine residues (75;98-100). These glycosylation patterns appear to be similar between different proteins and complexes with appropriate species specific difference in glycosylation. The exact role of glycosylation in cellulosomal linkers is unclear. Studies with designer recombinant cellulosomes produced in E. coli indicate that glycosylation is not required for activity (42;83;101). However, some speculate that heavy glycosylation may prevent proteolytic cleavage of the linkers or maybe provides additional structural stability required to link the cell to cellulose-substrates. Glycosylation may also affect linker conformational possibilities involved in substrate targeting. Nevertheless, the contribution of the linkers to cellulosome structure and function remains to be fully addressed.

### 6.6 Enzymes

In *C. thermocellum* alone, more than 70 different type-I dockerin containing enzymatic subunits have been identified based on sequence. These subunits are all theoretically capable of binding to any of the type-I cohesin modules of the cellulosomal
scaffoldin proteins. Recently, the type-I dockerin dual binding mode was discovered, which has added more plasticity to enzyme configuration on the scaffoldin (35;36). The number of possible combinations of enzymes, positions, and binding modes could yield a staggering number of possible cellulosome complexes. However, it is not known why there are so many possible complexes or how many of these complexes exist in nature.

The composition of the enzyme pool present to populate the cellulosome could be controlled through selective gene expression in response to environmental cues. Perhaps, some enzymes function better under different environmental conditions or on different substrates. As a result, gene expression is modulated accordingly. However, even with a restricted enzyme pool the number of enzymatic configurations is large. The presence of the enzymes in particular ratios may be sufficient to ensure maximal cellulolytic activity of the complex or maybe interactions occur between enzymes that help to arrange them specifically on the scaffoldin. SAXS studies described in this thesis demonstrate that contacts between neighbouring enzymes occur when the intervening linkers between two type-I cohesins are in a compact configuration. Finally, the dual binding mode of type-I dockerins may also be controlled by environmental factors or interactions with adjacent enzymes. Ultimately, the role of enzyme composition, arrangement, and binding plasticity remains to be addressed; however, several limitations on the study of the cellulosomes that are discussed below need to be overcome first.

6.7 Limitations of In vitro and In vivo Cellulosome Model Systems

Cellulosomes are produced by a variety of different anaerobic microbes. Several of these organisms have had their genomes sequenced. However, none have established
systems for genetic manipulation. Many of the questions that remain cannot easily be addressed without the ability to alter gene expression and/or sequence. For instance, why are there so many cellulosomal enzymes and anchoring proteins? Are all of the enzymes and anchoring proteins required? What is the role do posttranslational modifications play in cellulosome structure and function? Some studies have cultured cellulosome producing bacteria and either cleaved cellulosomes from the surface of the cell or employed secreted cellulosome systems for in vitro study (42;102-104). However, endogenous cellulosomes are highly heterogeneous both in terms of posttranslational modifications as well as enzyme position, orientation, and composition. In order to reduce this heterogeneity, in vitro studies of cellulosome structure and function are often performed using recombinant cellulosome components (42;83;101). These recombinant proteins lack all posttranslational modifications, which reduces heterogeneity but may affect protein structure and function. Furthermore, despite significant efforts, full-length recombinant cellulosomes cannot be produced. Most conclusions about cellulosome structure and function have been established by either designer cellulosomes or recombinant fragments of cellulosomal complexes, both of which may not represent endogenous systems very well. Finally, the substrate of cellulosomes is insoluble. Consequently, traditional methods for studying soluble enzyme-substrate reactions in solution cannot be applied and novel approaches need to be developed to apply to cellulosome systems as they become available.
6.8 Combining Techniques and the Dissect and Build Approach

High resolution structures of intact proteins and complexes at each stage of their mechanistic states is often not possible for a variety of reasons including poorly behaved samples, large size, and inherent flexibilities. There are some examples where such proteins have been crystallized and yielded high resolution structures such as the ribosome, the nucleosome and nucleosome arrays, several intact viral particles, and the two high resolution crystal structures from the cellulosome that are described in this thesis. However, there are also many examples of proteins and complexes that have failed to crystallize. With the growth in the application of high throughput technologies in structural biology, proteins that are amenable to structural study are being rapidly solved by structural genomics initiatives. Thus, only those proteins and complexes that are not well suited for traditional high resolution structure determination will remain poorly studied. In this thesis, we describe the use of a “dissect and build” approach to studying large proteins and complexes. The dissection stage has proven valuable in studying the cellulosome and other modular or multicomponent systems; however, the build stage is often more challenging. The major limitation is generating protein samples that are amenable to high resolution techniques. Since each structural technique has unique advantages and disadvantages, a combination of techniques can help to overcome the weaknesses of a given technique. In this thesis, we describe a successful example of the ‘dissect and build’ approach that combines high resolution structures with SAXS. It is likely that a combination of techniques with the ‘dissect and build approach’ will become
more common place in the future as structural biologists begin to tackle more challenging problems.
References


22. Schaeffer, F., Matuschek, M., Guglielmi, G., Miras, I., Alzari, P. M., Beguin, P. (2002) Duplicated dockerin subdomains of *Clostridium thermocellum* endoglucanase CelD bind to a cohesin domain of the scaffolding protein CipA
with distinct thermodynamic parameters and a negative cooperativity.  

*Biochemistry* **41**, 2106-2114


glycoproteins from the cellulosomes of *Clostridium thermocellum* and *Bacteroides cellulosolvens*. *Journal of Biological Chemistry* **268**, 26956-26960


to 1.8 angstrom resolution: insights into a new family of kinase structures. 

*Structure* 6, 183-193


Appendix I

**ADP-Dependent 6-Phosphofructokinase from *Pyrococcus horikoshii OT3*: Structure Determination and Biochemical Characterization of **PH1645**

A.1.1 Abstract

Some hyperthermophilic archaean use a modified glycolytic pathway that employs an ADP-dependent glucokinase (ADP-GK) and an ADP-dependent phosphofructokinase (ADP-PFK) or, in the case of *Methanococcus jannaschii*, a bifunctional ADP-dependent gluco-phosphofructokinase (ADP-GK/PFK). The crystal structures of three ADP-GKs have been determined. However, there is no structural information available for ADP-PFKs or the ADP-GK/PFK. Here, we present the first crystal structure of an ADP-PFK from *Pyrococcus horikoshii OT3 (PhPFK)* in both apo and AMP-bound forms determined to 2.0 Å and 1.9 Å resolution, respectively, along with biochemical characterization of the enzyme. The overall structure of PhPFK maintains a similar large and small α/β domain structure seen in the ADP-GK structures. A large conformational change accompanies binding of phosphoryl donor, acceptor, or both, in all members of the ribokinase superfamily characterized thus far, which is believed to be critical to enzyme function. Surprisingly, no such conformational change was observed in the AMP-bound PhPFK structure compared to the apo structure. Through comprehensive site-directed mutagenesis of the substrate binding pocket we identified residues that were critical for both substrate recognition and the phosphotransfer reaction. The catalytic residues and many of the substrate binding residues are conserved between PhPFK and
ADP-GKs; however, four key residues differ in the sugar binding pocket, which we have shown determine the sugar binding specificity. Using these results we were able to engineer a mutant PhPFK that mimics the ADP-GK/PFK and is able to phosphorylate both fructose-6-phosphate and glucose.

A.1.2 Introduction

Glycolysis is a central and highly conserved metabolic pathway in all three domains of life. However, the Embden-Meyerhof glycolytic pathway of some hyperthermophilic archaea displays distinct differences from the classical pathway. Glyceraldehyde-3-phosphate is converted to 3-phospho (3P)-glycerate by glyceraldehyde-3-phosphate ferredoxin oxidoreductase in a single-step instead of using glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate kinase in the classical two-step reaction mechanism (105,106). Also, the classical ATP-dependent glucokinases (ATP-GKs) and phosphofructokinases (ATP-PFKs) are replaced with novel ADP-dependent glucokinases and ADP-dependent phosphofructokinases or, in the case of Methanococcus jannaschii, a bifunctional ADP-dependent gluco-/phosphofructokinase (ADP-GK/PFK) (107-111).

PFKs convert fructose 6-phosphate (F6P) to fructose 1,6-bisphosphate. This is an early step in the EM pathway and therefore represents a critical control point for the entire pathway. Sequence and structural characterization classify PFKs into two convergent protein families (112): the PFK-A family and the ribokinase superfamily, which includes the PFK-B family and ADP-GK/ADP-PFK family. The PFK-A family
includes both allosterically regulated ATP-dependent enzymes found in a variety of eukarya and bacteria and non-allosterically regulated PPI-dependent enzymes found in all three domains of life (113;114). The PFK-B family is a diverse family of ATP-dependent carbohydrate and pyrimidine kinases that is also present in all three domains of life. In general, ATP-PFKs from the PFK-B family can be differentiated from PFK-A ATP-PFKs by their lack of allosteric regulation, although the minor ATP-PFK from *Escherichia coli*, Pfk-2, is allosterically regulated by MgATP (115).

Several crystal structures of PFK-A family PFKs have been reported. The best studied are the tetrameric PFKs from *E. coli* and *Bacillus stearothermophilus* (116;117). Both contain subunits that consist of a large and a small 3-layered αβα sandwich domain (116;117). The large domain binds to ATP and the small domain binds to F6P (116;117). Recently the crystal structure of Pfk-2, a member of the PFK-B family, was solved (118). It is composed of two domains, a large 3-layered αβα sandwich domain, similar to that seen in the PFK-A family protein structure mentioned above, and a smaller four-stranded β-sheet domain (118).

To date ADP-PFKs have been reported as non-regulated enzymes with Michaelis-Menten kinetics. This part of glycolysis is particularly interesting since with the presence of PFKs and fructose-1,6bisphosphatase (FBPase) there is the possibility to produce futile cycling resulting in a net hydrolysis of nucleotide (ADP in the case of thermococcales) which, of course is undesirable since it would uncouple the metabolism. *E. coli* has overcome this problem by using a strong MgATP induced inhibition when F6P is low in both of its PFKs: Pfk-1 from PFK-A family and Pfk-2 from the ribokinase superfamily.
On the other hand, the archaea with the modified Embden-Meyerhof pathway seems to control this issue at the transcriptional level as, depending on the growth medium, either the production of PFK or FBPase is turned off (121). In this way, regulation of PFK is not needed.

The crystal structures of the ADP-GK from *P. horikoshii* (*PhGK*), the ADP-GK from *Thermococcus litoralis* (*TlGK*) bound to ADP, and the ADP-GK from *Pyrococcus furiosus* (*PfGK*) in complex with AMP and glucose have been reported, although there are no crystal structures of ADP-PFK or the bifunctional ADP-GK/PFK (122-124). Despite little sequence identity between ADP-GKs and other kinases the fold is similar to that of ribokinases (122). As a result, ADP-GKs have been classified as members of the ribokinase superfamily (122). Overall, ADP-GK structures are composed of a large and a small α/β domain (122-124). The ADP binding pocket is found on the surface of the large domain (122-124). Residues responsible for glucose binding are found on the surface of both the large and small domains adjacent to the ADP-binding pocket (122-124). Upon glucose binding, the protein undergoes a conformational change whereby the large and small domains close in creating the active-site pocket (123). Interestingly, the unique ADP binding pocket of ADP-GKs recognizes the α- and β-phosphate of the ADP in an almost identical manner to how the β- and γ-phosphate of ATP are recognized in ATP-dependent kinases of the ribokinase superfamily (122).

Here, we report the first structure of an ADP-PFK, alone and in complex with AMP, together with the biochemical characterization of the kinetic properties and substrate specificities. Moreover, we carried out site-directed mutagenesis in the sugar
binding pocket and active site of *Ph*PFK and identified residues that are critical for PFK activity and for distinguishing between glucose and F6P.

**A.1.3 Experimental Procedures**

**A.1.3.1 Cloning, Expression, and Protein Purification**

The open reading frame for *Ph*PFK (*PH1645*) was amplified by PCR and inserted between the NdeI and BamHI restriction sites of a modified pET-15b expression vector (Novagen) (vector p11) as previously described (125). This construct generated an N-terminal hexahistidine tag joined to the *Ph*PFK protein by the TEV protease recognition site (ENLYFQ↓G). Recombinant wild type and mutant (see below) native and selenomethionine labeled *Ph*PFK proteins were expressed in BL21 (DE3) and DL41 (DE3) E. coli cells (Novagen), respectively. Protein was purified using Ni²⁺-NTA affinity and size exclusion chromatographies. For more details see supplemental materials.

**A.1.3.2 Site-Directed Mutagenesis**

Site-directed mutagenesis was performed using QuickChange™ site-directed mutagenesis kit (Stratagene) according to the manufacturer’s protocol using the *Ph*PFK expression vector described above as the template. See Table A.1-1 for list of mutations made.

**A.1.3.3 Protein Crystallization**

Protein crystals of apo-*Ph*PFK were generated through hanging drop vapor diffusion at 21 °C by mixing 2 µl of protein solution (10mg/mL) with 2 µl of well
Table A.1-1 Kinetic parameters of wild type and mutant PhPFKs.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_{cat}$</th>
<th>$K_M$ F6P</th>
<th>$k_{cat} / K_M$ F6P</th>
<th>$K_M$ ADP</th>
<th>$k_{cat} / K_M$ ADP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s$^{-1}$</td>
<td>$\mu M$</td>
<td>M$^+ s^{-1}$</td>
<td>$\mu M$</td>
<td>M$^+ s^{-1}$</td>
</tr>
<tr>
<td>Wild</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>45.5 ± 4.0</td>
<td>15.2 ± 2.5</td>
<td>2.99 • 10$^6$</td>
<td>13.2 ± 2.4</td>
<td>3.45 • 10$^6$</td>
</tr>
<tr>
<td>N15A</td>
<td>91.1 ± 3.9</td>
<td>103 ± 1.4</td>
<td>8.84 • 10$^5$</td>
<td>23.8 ± 2.6</td>
<td>3.83 • 10$^6$</td>
</tr>
<tr>
<td>D17A</td>
<td>1.98 ± 0.0051</td>
<td>22600 ± 840</td>
<td>8.76 • 10$^1$</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>A71E</td>
<td>39.3 ± 9.9</td>
<td>3.95 ± 0.2</td>
<td>1.72 • 10$^6$</td>
<td>13.6 ± 0.6</td>
<td>2.82 • 10$^6$</td>
</tr>
<tr>
<td>Q97A</td>
<td>13.0 ± 1.0</td>
<td>7.08 ± 0.74</td>
<td>1.84 • 10$^6$</td>
<td>7.47 ± 0.70</td>
<td>1.74 • 10$^6$</td>
</tr>
<tr>
<td>K158A</td>
<td>41.0 ± 6.7</td>
<td>6500 ± 1300</td>
<td>6.31 • 10$^3$</td>
<td>37.9 ± 6.7</td>
<td>1.08 • 10$^6$</td>
</tr>
<tr>
<td>N160A</td>
<td>151 ± 16</td>
<td>415 ± 13</td>
<td>3.64 • 10$^5$</td>
<td>72.4 ± 4.9</td>
<td>2.09 • 10$^6$</td>
</tr>
<tr>
<td>S183E</td>
<td>91.7 ± 23.6</td>
<td>86.5 ± 9</td>
<td>1.06 • 10$^6$</td>
<td>29.6 ± 2.2</td>
<td>3.10 • 10$^6$</td>
</tr>
<tr>
<td>R185A</td>
<td>0.351 ± 0.016</td>
<td>318 ± 9.5</td>
<td>1.10 • 10$^3$</td>
<td>473 ± 24</td>
<td>7.42 • 10$^2$</td>
</tr>
<tr>
<td>S189A</td>
<td>25.5 ± 2.6</td>
<td>9.02 ± 2.3</td>
<td>2.83 • 10$^6$</td>
<td>16.0 ± 1.2</td>
<td>1.60 • 10$^6$</td>
</tr>
<tr>
<td>R191A</td>
<td>27.4 ± 1.5</td>
<td>254.4 ± 26.1</td>
<td>1.07 • 10$^5$</td>
<td>14.5 ± 1.2</td>
<td>1.89 • 10$^6$</td>
</tr>
<tr>
<td>R191E</td>
<td>42.5 ± 1</td>
<td>4870 ± 173</td>
<td>8.73 • 10$^3$</td>
<td>60.5 ± 3.2</td>
<td>7.02 • 10$^5$</td>
</tr>
<tr>
<td>Q224A</td>
<td>65.3 ± 2.6</td>
<td>38.6 ± 1.3</td>
<td>1.69 • 10$^6$</td>
<td>29.5 ± 5.0</td>
<td>2.21 • 10$^6$</td>
</tr>
<tr>
<td>S263A</td>
<td>64.1 ± 14</td>
<td>39.5 ± 7.8</td>
<td>1.62 • 10$^6$</td>
<td>20.8 ± 3.4</td>
<td>3.08 • 10$^6$</td>
</tr>
<tr>
<td>D433A</td>
<td>0.00480 ± 0.00040</td>
<td>12.7 ± 4.2</td>
<td>3.78 • 10$^2$</td>
<td>13.9 ± 3.3</td>
<td>3.45 • 10$^2$</td>
</tr>
</tbody>
</table>

n.d.: not determined. For the D17A mutant the kinetic parameters for MgADP were not determined since the elevated $K_M$ value for F6P hinders us to reach saturating conditions.
solution consisting of 22 % PEG 4000, 0.1 M Tris-HCl, pH 8.5, and 0.2 M LiSO4. The crystals of *Ph*PFK complex with AMP were obtained by crystallization of the *Ph*PFK D17A protein in the presence of 20% PEG 3350, 0.2 M Li-citrate, 10 mM fructose 6-phosphate, and 5 mM ADP. Prior to data collection, crystals were transferred into Paratone-N and cryo-cooled in a nitrogen-gas steam.

**A.1.3.4 X-ray Diffraction, Structure Determination, and Refinement**

Apo-*Ph*PFK crystals were placed in a cryoprotectant composed of 15% glycerol added to the crystallization solution and then flash frozen in liquid nitrogen prior to data collection. Multiwavelength anomalous dispersion (MAD) data was collected at three wavelengths 0.96396, 0.97918, 0.97943 Å at the Advanced Photon Source (APS, Argonne, IL, USA) beamline 19-ID of the Structural Biology Center-CAT with a SBC-3 CCD detector. The data was processed using HKL2000 (70). Data collection and processing statistics are shown in Table A.1-2.

The structure of *Ph*PFK was solved using the MAD method. Selenium sites were located using SOLVE (126;127). Six of the 10 expected selenium sites per asymmetric unit were found. Selenium position refinement, phase calculation and density modification was performed by SHARP (128). The structural model was built and refined by XFIT, and CNS and Refmac (129;130). The AMP-bound structure was solved by molecular replacement using apo-*Ph*PFK structure. The final refinement statistics can be found in Table A.1-2.
Table A.1-2 Crystallographic data and refinement statistics.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Apo-PhPFK</th>
<th>Complex with AMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁</td>
<td>P2₁</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td>(a=68.07, b=99.93, c=82.58 \text{ Å}, \beta=110.38^\circ)</td>
<td>(a=68.5, b=104.2, c=70.8 \text{ Å}, \beta=105.1^\circ)</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>40.0 - 2.0</td>
<td>41.5-1.9</td>
</tr>
<tr>
<td>Unique/Free Reflections (5%)</td>
<td>69682/3508</td>
<td>71005/3770</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.2% (98.6%)</td>
<td>98.7% (90.7%)</td>
</tr>
<tr>
<td>(R_{\text{merge}})</td>
<td>0.08 (0.445)</td>
<td>0.08 (0.353)</td>
</tr>
<tr>
<td>(I/\sigma I)</td>
<td>21.0 (3.57)</td>
<td>16.7 (3.25)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.6 (5.9)</td>
<td>4.5 (3.9)</td>
</tr>
<tr>
<td>(R/R_{\text{free}}) (%)</td>
<td>0.202/0.245</td>
<td>17.3/22.1</td>
</tr>
<tr>
<td>r.m.s.d. bond lengths (Å)</td>
<td>0.004</td>
<td>0.005</td>
</tr>
<tr>
<td>r.m.s.d. bond angles (°)</td>
<td>0.609</td>
<td>0.705</td>
</tr>
<tr>
<td>Mean temperature factor (Å²)</td>
<td>27.65</td>
<td>29.51</td>
</tr>
</tbody>
</table>

\(R_{\text{sym}} = \frac{\Sigma|I(k)| - \langle I \rangle}{\Sigma I(k)},\) where \(I(k)\) and \(\langle I \rangle\) represent the diffraction intensity values of the individual measurements and the corresponding mean values. The summation is over all measurements. Values in parentheses are for the highest resolution shell. PDB codes are 1U2X and 3DRW for apo-PhPFK and its complex with AMP, respectively.
A.1.3.5 Determination of Enzyme Activity
Phosphofructokinase (PFK) activity was assayed spectrophotometrically at 50 ºC, by coupling the fructose-1,6-bisphosphate formation to the oxidation of NADH as previously described (107). See Supplemental Materials for more details.

A.1.3.6 Substrate Specificity
The phosphoryl group donor specificity of PhPFK was determined by measuring the PFK activity as described above substituting ADP for the following phosphoryl group donors, UDP, IDP, GDP, and CDP. The activity is reported as percentage of the maximum activity obtained with ADP. The divalent cation preference was tested similarly, replacing MgCl₂ in the assay by MgSO₄, MnCl₂, NiSO₄, CaCl₂, or ZnCl₂. 2.5, 4 and 7 mM total metal were used, keeping other components as those used in the standard assay. The activity observed using these metals was reported as U/mg.

A.1.3.7 Kinetic Parameters
Kinetic parameters were determined at 50º C by varying the concentrations of F6P (0 to 2 mM) with saturating ADP concentrations, and vice verse (104). For some mutants, the range of F6P concentrations was increased. In all these experiments, MgCl₂ was used as the divalent cation. The data was analyzed using SigmaPlot Software (Systat Software, Inc. CA, USA), and fitted to the Michaelis-Menten equation.

A.1.4 Results and Discussion
A.1.4.1 Enzymatic Activity and Substrate Specificity of PhPFK

PhPFK is predicted to be an ADP-PFK based on sequence similarity (Figure A.1-1). In addition, the PhPFK structures align well with the available ADP-GK structures (Figures A.1-2C and A.1-2D). Therefore, to test whether this enzyme is a true ADP-PFK the phosphorylation of F6P was assayed in the presence of several nucleotides. PFK activity was the highest in the presence of ADP or UDP as the phosphoryl group donor (Figure A.1-3A). However, PhPFK is also capable of transferring phosphoryl groups from IDP, GDP, and CDP to F6P but at a significantly slower rate (Figure A.1-3A). A saturation curve was generated for both ADP and UDP. The calculated $k_{cat} K_M$ value for UDP was 40 times higher than that of ADP, indicating a strong preference for ADP over UDP as the phosphoryl group donor.

Wild type PhPFK displayed Michaelis-Menten kinetics at 50°C. The following constants were determined using direct fit: $K_M$ values of 15.2 ± 2.5 and 13.2 ± 2.4 µM for F6P and ADP, respectively, and a $k_{cat}$ value of 45.5 ± 4.0 s⁻¹. All kinetic parameters are summarized in Table A.1-1 and Figure A.1-3B. The temperature of these reactions is much lower than the optimal growth temperature of P. horikoshii and since temperature has a significant effect on enzyme kinetics, the constants calculated here may differ from the values at the optimal growth temperature. Unfortunately, the commercially available auxiliar enzymes for this reaction are from mesophilic sources hindering the use of high temperatures in the assay. On the other hand, due to the large amount of mutants characterized in this article the study would be prohibitively longer using a discontinuous assay to measure the kinetic parameters at higher temperatures. Nevertheless, most of the
Figure A.1-1 Multiple sequence alignment of \textit{PhPFK} with other ADP-PFKs, ADP-GKs, and the bifunctional ADP-GK/PFK. (\textit{Pf}, \textit{Pyrococcus furiosus}; \textit{Ph}, \textit{Pyrococcus horikoshii}; \textit{Tl}, \textit{Thermococcus litoralis}; and \textit{Mj}, \textit{Methanococcus jannaschii}). This figure was prepared using GeneDoc (131).
Figure A.1-2 Structure of apo and AMP-bound PhPFK and comparisons with ADP-GKs. A. Apo-PhPFK. B. AMP-bound PhPFK. C. Alignment of apo-PhPFK structure (green) with PhGK structure (cyan). The additional N-terminal α-helix of PhGK not found in PhPFK is shown in magenta. D. Alignment of AMP bound PhPFK structure (green) and ADP bound TlGK structure (cyan). The additional N-terminal α-helix of TlGK not found in PhPFK is shown in magenta. This figure was generated using PyMOL.
A

maximal activity (%)

Wild type
S183E mutant

ADP  IDP  UDP  GDP  CDP  Control
Figure A.1-3 PhPFK phosphoryl donor, kinetics, and divalent cation specificity. A. Wild type and S183E PhPFK were assayed with different nucleotides as phosphoryl donors. The activity is shown as the percentage of the activity measured in presence of ADP (100%). B. Phosphofructokinase activity measurements were made for wild and mutant PhPFKs. C. PhPFK activity measured in presence of 2 mM F6P, 2 mM ADP, and the indicated divalent cation concentration. “Control” corresponds to the activity in the absence of nucleotide (A) and metal (C). Results are given as the mean ± S. E. of three experiments.
work published for other enzymes from this family uses the same strategy, making the data here directly comparable. The calculated $K_M$ values for F6P and ADP for wild type $Ph$PFK are significantly lower than those of the ADP-PFK from $P. furiosus$, 2.3 and 0.11 mM, respectively (111). However, the $K_M$ for F6P is similar to the $K_M$ obtained for the bifunctional ADP-GK/PFK from $M. jannaschii$, 9.6 µM. But, the $K_M$ for ADP is still much lower, 0.49 mM (110).

Divalent cations were required for $Ph$PFK activity. Five divalent metal cations were tested (Magnesium was tested with two counter ions to discard any effect of the anion) using 2.5, 4 and 7 mM of total metal with 2 mM total ADP and 2 mM total F6P (Figure A.1-3C). In these conditions, the highest $Ph$PFK activity was obtained with NiSO$_4$ (~ 97 U/mg). The enzyme was also active in the presence of MgCl$_2$, MnCl$_2$, MgSO$_4$, and CaCl$_2$, but no significant activity was detected when the assay was performed in the presence of ZnCl$_2$. An increase in the enzyme activity, concomitantly with the increase in MnCl$_2$ concentration was observed, while the opposite effect is observed in the presence of NiSO$_4$. For the other divalent metals, the activity did not change with the cation concentration in the range tested. CaCl$_2$ showed the lowest activity with ~75% of the activity measured in the presence of MgCl$_2$. Although it is generally accepted that magnesium is the in vivo preferred metal, the ability of Ni$^{2+}$, Mn$^{2+}$, and to a lower extent Ca$^{2+}$ to support $Ph$PFK activity with a high catalytic rate, suggests that other metals can substitute Mg$^{2+}$. 

162
A.1.4.2 Overall Structure

The apo-PhPFK structure (Figure A.1-2A) was refined to 2.0 Å resolution with R = 0.177 and R<sub>free</sub> = 0.226 (Table A.1-2). The structure contained 450 of the expected 474 residues from the expressed protein construct (Table A.1-2). In the final model, 97.9% of the residues were in favored regions of the Ramachandran plot and none were in the disallowed region (Table A.1-2). PhPFK was crystallized alone and in complex with ADP. The apo structure was solved by MAD method using a selenomethionine derivative and the complex structure was solved by molecular replacement using the apo structure as a model. AMP, instead of ADP, was found in the active site. The AMP-bound structure (Figure A.1-2B) was refined to a resolution of 1.9 Å with R = 0.177 and R<sub>free</sub> = 0.228 (Table A.1-2). Both the apo and AMP-bound PhPFK crystal structures contained two molecules in the asymmetric unit. For a more detailed discussion of oligomeric state see supplemental materials.

The crystal structure of PhPFK was aligned with PhGK using the DaliLite server (Figure A.1-2C) (85). PhPFK and PhGK share 31% identity (Figure A.1-1). Overall, 407 residues were aligned, resulting in a Z-Score and RMSD of 39.9 and 2.5 Å, respectively. The AMP bound PhPFK structure was aligned with the ADP bound TlGK structure using the DaliLite server (Figure A.1-2D) (85). These two proteins also shared 31% identity with one another (Figure A.1-1). A total of 434 residues were aligned with a Z-Score of 41.1 and a RMSD of 3.3 Å. In both cases the structures were similar, sharing the same large and small α/β domain structures suggesting the common evolutionary origin of ADP-PFKs and ADP-GKs. However, the ADP-GK structures have an additional α-helix
at their N-termini and in the ADP bound Tl/GK structure, the small domain is closer to the large domain compared to AMP bound PhPFK. This is discussed further below.

A.1.4.3 Nucleotide and Mg$^{2+}$ Binding Site

Overall, the apo and AMP bound PhPFK structures show very little difference (RMSD = 0.7 Å; Z-Score = 61.9). However, AMP binding results in slight changes in the large nucleotide binding loop as well as the main chain positions of the β-sheet and loop that lie directly N-terminal to the large nucleotide binding loop (Figure A.1-4). A conformational change of the large nucleotide binding loop was also reported upon ADP binding to Tl/GK. There is good conservation between the nucleotide binding site of PhPFK and ADP-GKs, however some differences exist in the sugar bound, closed conformation (threading model, see below). In the AMP and glucose bound PfGK structure, Y344 of the large domain hydrogen bonds with E195 from the small domain. However, in PhPFK the equivalent small domain residue, S183, is too short to hydrogen bond with Y338 of the large domain. This difference may account for the difference in phosphoryl group donor specificity in ADP-GKs and ADP-PFKs. ADP-PFKs are able to use GDP, to some extent, whereas ADP-GKs cannot. We made a S183E mutation and tested its ability to use different phosphoryl donors compared to the wild type enzyme (Figure A.1-3A). There was no difference in activity between the mutant and wild type using ADP as a phosphoryl donor. However, the S183E mutant displayed significantly less activity when the phosphoryl donor was replaced with IDP, UDP, GDP, or CDP, compared to the wild type.
Figure A.1-4 AMP binding site of PhPFK. A. Stereo view of AMP-bound PhPFK nucleotide binding site. B. Comparison of the nucleotide binding site of apo (green) and AMP-bound (cyan) PhPFK. These figures were generated in PyMOL.
One of the most intriguing features of these enzymes is the position in that they can bind ADP in a way that both α- and β-phosphates are recognized mainly in the same way as the β- and γ-phosphates in the ATP dependent enzymes. In this respect, it has been proposed that the presence of a bulky side chain in the bottom of the nucleotide binding crevice (Y357 in the case of *T. litoralis* ADP-GK), below the sugar part of the molecule, could account for the change in specificity (123). Moreover, it was proposed that since ADP-PFKs have leucine or isoleucine in this position they can marginally use ATP as phosphoryl donor as it was seen for the enzyme from *P. furiosus* (111). Since the corresponding lateral chain of *Ph* PFK (L340) is placed in the corresponding position we tested directly the ability of this enzyme to use ATP using the Pyruvate kinase/Lactate dehydrogenase coupled assay, which specifically consumes the ADP produced. Surprisingly, no activity was detected with this nucleotide.

Almost all phosphate transferring enzymes have Mg$^{2+}$ in their active site thereby assisting with the phosphate transfer through interactions with the β- and γ-phosphate groups of ATP or the β-phosphate of ADP, when ADP is the phosphoryl group donor. However, Mg$^{2+}$ was not observed in the *Ph* PFK structure or any of the ADP-GK structures. Although, 17 water molecules were trapped in the AMP and glucose bound *Pf* GK structure, which offers sufficient space for both a magnesium ion and the β-phosphate of ADP. Despite the prerequisite for Mg$^{2+}$, this cation has never been observed in the structure of a member of this superfamily, including *Ph* PFK.
A.1.4.4 Reaction Mechanism

A conserved aspartate residue in ribokinases, adenosine kinases, ADP-GKs, ADP-PFKs, and ADP-GK/PFK is believed to be the catalytic base involved in the phosphotransfer reaction of these kinases (Figure A.1-1). Mutations of this residue in TlGK resulted in a significant reduction of glucokinase activity (123). Similarly, mutation of the corresponding aspartate in PhPFK (D433) to alanine resulted in a loss of PFK activity (Table A.1-1).

Furthermore, a conserved arginine residue that approaches the active site during the domain closing event associated with sugar binding is believed to attract the terminal phosphate group of the phosphoryl donor, stimulating cleavage of the phosphodiester bond and transfer of the phosphate group. Mutation of R205 in TlGK to alanine resulted in <0.1% the activity of the wild type enzyme (123). We observed the same effect upon mutation of the corresponding arginine residue in PhPFK (R185) to alanine (Table A.1-1). These two findings support a conserved mechanism amongst ADP-GKs, ADP-PFKs, and ADP-GK/PFKs.

A.1.4.5 Substrate Induced Fit

A large conformational change occurs in ADP-GKs as well as ATP-dependent hexokinases/glucokinases and ribokinases upon binding their respective substrates (Figure A.1-5) (123;132-135). The flexible loops located between the large and small domains form a hinge that folds, closing the cleft between the two domains upon substrate binding (123;132-135). This conformational change appears to be essential for positioning the catalytic residues of ADP-GKs and ribokinases (123;134). The three
Figure A.1-5 Substrate induced conformational changes of ADP-GKs. Left, apo structure from *Pyrococcus horikoshii* (124); middle, ADP-bound structure from *Thermococcus litoralis* (122); and right, AMP and glucose bound structure from *Pyrococcus furiosus* (123).
ADP-GKs structures all display varying amounts of domain closure (Figure A.1-5) (122-124). The apo-PhGK is the most open conformation (Figure A.1-5) (124). There is a rotation of about 20º between the large and small domain in the ADP bound TlGK structure compared to the PhGK structure and an additional 20º rotation between the two domains in the AMP and glucose bound PfGK structure compared to the TlGK structure (Figure A.1-5) (122-124). It was initially reported that the TlGK structure was the most open conformation because they were able to successfully soak ADP out of the crystal, which indicates the absence of a large conformational change (122). The PhGK structure clearly presented a more open conformation, which led to the suggestion that TlGK had been trapped in the ADP bound conformation following the removal of the ADP from the crystal (124). However, ADP could not be soaked into the PhGK crystal, which is thought to be due to an ADP-dependent domain closing (124). Interestingly, no such conformational change was observed between the apo and AMP-bound PhPFK structures. It could be that the β-phosphate of ADP is required to fully induce this conformational change. The amine nitrogen of R205 from TlGK is 3.8 Å away from the β-phosphate of ADP and may form important interactions that help to stabilize a more closed conformation (122). One other possibility is that for PhPFK both co-factor and substrate are required for the conformational change observed as seen in other structures. In any case, the lack of conformational change in PhPFK in the presence of AMP is somewhat surprising given the structure is from co-crystallization instead of soaking.
A.1.4.6 Sugar Binding Site

Attempts to co-crystallize or soak crystals with F6P or fructose-1, 6-bisphosphate both in the presence or absence of ADP or nonhydrolyzable ADP analogues were unsuccessful. Due to the overall similarity between the PhPFK structures and ADP-GKs structures, we modeled PhPFK in the closed conformation bound to F6P and AMP based on the AMP and glucose bound PfGK structure (Figure A.1-6). The primary sequence of PhPFK was threaded onto the PfGK structure bound to both AMP and glucose, and the resulting model was energy minimized. F6P was docked with the structure using AutoDock (136). All of the ten lowest energy docking results placed the phosphate in the same position with only slight differences in the placement of the sugar moiety.

To validate our model and gain further insights into substrate specificity, relevant residues from the F6P binding pocket of PhPFK, the highly conserved residues N15, D17, Q97, K158, N160, R185, S189, R191, Q224, S263, and D443, were mutated to alanine in order to establish their importance to substrate binding and catalysis. Of these, N15A, D17A, K158A, N160A, R185A, and R191A show the most significant increase in their K_M value for F6P (Table A.1-1). N160, N15, and R191 form one, two, and three hydrogen bonds with the phosphate group of F6P, respectively as judged by the docked conformation (Figure A.1-6). R185 and D17 form hydrogen bonds with the sugar moiety of F6P (Figure A.1-6). As a result, loss of these interactions weakens PhPFK binding to its substrate. K158 is too far to hydrogen bond with the phosphate of F6P in our model (Figure A.1-6). However, it is strictly conserved in PFKs and mutation of this residue has the most significant effect on F6P binding after the D17A mutation (Table A.1-1).
Figure A.1-6 Model of PhPFK bound to F6P and AMP. Docking Model of the PhPFK-F6P Complex. Where different, the positions of residues conserved in ADP-GKs are shown in orange. The red residue corresponds to the glutamate residue conserved in the bifunctional ADP-GK/PFK (E82) and in ADP-GKs (E88). Black hashed lines represent hydrogen bonding.
are several possible explanations for this observation; however, it is most likely a result of the threading procedure itself, since there is no negative charge to stabilize the positive lateral chain of K158 inside the binding pocket and therefore repelling it in the opposite direction. In fact, simple rotation of the side chain of K158 in our model would result in a distance of less than 3.5 Å to the phosphate group of F6P. This was confirmed using a molecular dynamics simulation (see Supplemental Materials). In this conformation, the lateral chains of K158 and R191 could form a salt bridge with the phosphate group of F6P, which would explain the behavior of the mutant enzyme.

Interestingly, the half of the sugar ring opposite to the endocyclic oxygen is recognized mainly in the same way as the comparable part of glucose by ADP-GKs. As expected the residues responsible of this like N15, D17, and G96 are conserved in both specificities. As can be seen below the sugar discrimination comes mainly by the interactions with the other half of the ring. This suggests that the sugar binding site in the family is strongly plastic to accommodate a furanose-P and a pyranose with mainly the same residues, but still very selective to precisely discriminate between the two of them. Q224A and S263A contribute little to the sugar binding: both mutations result in small increases in $K_M$ for F6P of about 2- to 3-fold (Table A.1-1). Surprisingly, Q97A, S189A, and D433A mutations lead to a small decrease in $K_M$ for F6P despite the loss of a hydrogen bond in the case of Q97A and D433A (Table A.1-1). Also, R185 and D433 make contact with the C1 hydroxyl group, which supports the proposed catalytic mechanism. However, due to the lack of the phosphate group from ADP, the position of
R185 could be somewhat distorted. Nevertheless, our model is well supported by the mutational data.

**A.1.4.7 Differences Between ADP-PFKs and ADP-GKs**

In our model, the conserved residues N13, N15, D17, G96, Q97 and R185 form hydrogen bonds with F6P (Figure A.1-6). K158, N160, and R191 also form hydrogen bonds with F6P (Figure A.1-6). Both are present in the bifunctional ADP-GK/PFK from *M. jannaschii* but the equivalent residues in ADP-GKs are a non-conserved small residue, histidine and aspartate, respectively. A similar scenario was seen in a previous in silico study using the bifunctional enzyme from *M. jannaschii* (137). The positive charge on R191 of *Ph*PFK attracts the phosphate group of F6P, and is still compatible with glucose binding. However, when replaced by aspartate in ADP-GKs, the negatively charged side chain likely repels the phosphate group of F6P, which prevents ADP-GKs from binding F6P. In order to assess the role of R191 in F6P binding, we made two mutations, R191A and R191E. The R191A mutation removes three hydrogen bonds to the phosphate group on F6P and causes a ~17-fold increase in the $K_M$ value for F6P (Table A.1-1). However, as predicted, the charge reversal of the R191E mutation has a much more pronounced effect on F6P binding, a ~325-fold increase in $K_M$ (Table A.1-1). Even so, the R191E mutant enzyme is not able to phosphorylate glucose, suggesting that the presence of this charged residue in the bottom of the sugar binding site in all ADP-GKs contribute to sugar discrimination by hindering the protein from binding F6P and not by increasing the affinity for glucose.
The key difference between kinases that are capable of phosphorylating glucose and those that cannot is a conserved glutamate residue found in ADP-GKs (E88 from *PfGK*) and ADP-GK/PFK (E82) that is replaced with alanine in *PhPFK* (A71). In the *PfGK*-glucose-AMP complex structure, E88 lateral chain forms a hydrogen bond with the C2 hydroxyl group of the bound glucose molecule. This interaction is thought to be critical for orienting the glucose molecule in the active site and it has also been highlighted as crucial for the sugar specificity in the past (110;116;137). As a result, ADP-PFKs cannot phosphorylate glucose. We tested this hypothesis by mutating A71 to glutamate. As expected, this mutation has little effect on F6P binding compared to the wild type (Table A.1-1). However, the A71E mutant enzyme can also bind glucose ($K_M = 3.95 \pm 0.2$ mM) and is capable of phosphorylating glucose ($k_{cat}$ of $2.68 \pm 0.05$ s$^{-1}$) (Table A.1-1). In contrast, the wild type enzyme displays no activity for glucose. This result demonstrates that A71 is a key player in determining substrate specificity. Interestingly, the catalytic mechanism described before for the *PhPFK* seems to be even more general since comparison of these features with the non-homologous ATP dependent PFK (*Pfk*-1) from *E. coli* pointed out striking similarities.

In terms of catalysis, in Pfk-1, Asp127 plays a critical role as a general base, increasing the nucleophilicity of the 1-hydroxyl of F6P by abstracting its proton and permitting attack on the $\gamma$-phosphate of the substrate ATP in a similar way as D433 in *PhPFK* (138). In addition, R162 and R243 interact with the 6-phosphate of F6P. Truncation of these residues to serine results in enzymes with deceased F6P binding ability and reduced cooperativity, but little change in catalytic ability in a very similar
way to K158 and R191 (139). Then, it would appear that some structural features related to catalysis in ADP dependent PFK have analogous counterparts in the ATP-PFKs.

A.1.4.8 Conclusion

In summary, we determined the first crystal structure of an ADP-dependent phosphofructokinase both alone and in complex with AMP and through comprehensive mutagenesis we identified residues that were critical for both the phosphotransfer reaction and substrate binding. Our results demonstrate that the overall structure of ADP-PFK is similar to that of ADP-GK and that the two share a common mechanism of action. However, unlike the available ADP-GK structures, no conformational change was observed upon nucleotide binding. This was unexpected and may indicate that the nucleotide-dependent conformational change is a non-essential part of ADP-dependent sugar kinase mechanism. Moreover, we identified four key residues responsible for the sugar binding specificity. Using this knowledge, we were able to generate a mutant PhPFK that is capable of phosphorylating both F6P and glucose.

A.1.5 Footnotes/Acknowledgements

We thank all members of the Ontario Centre for Structural Proteomics (Structural Proteomics in Toronto, SPiT) for their help in conducting experiments. We also thank Dr. Richard C. Garratt from Instituto de Física de São Carlos, Universidade de São Paulo, Brasil, where the DLS experiments were performed. The work was supported by Genome Canada (through the Ontario Genomics Initiative), by National Institutes of Health grant GM074942, by grant from Fondo Nacional de Desarrollo Científico y Tecnológico
A.1.6 Supplemental Materials

A.1.6.1 Protein Expression and Purification

Cells were cultivated at 37°C in 1 L of Luria Bertani broth medium containing 100 µg/mL ampicillin and 35 µg/mL kanamycin until the OD$_{580}$ nm reached 0.5. Isopropyl-ß-D-thiogalactopyranoside was added to the medium to a final concentration of 1 mM to induce expression, and the cells were cultured for an additional 4 hours. Cells were harvested by centrifugation, resuspended in Buffer A (20 mM imidazole, 5 mM MgCl$_2$, 0.3 M NaCl and 20 mM NaH$_2$PO$_4$, pH 7.4), and disrupted by sonication. The crude extract was heated at 70°C for 30 min, and the denatured protein was then removed by centrifugation (3220 x G for 15 min). The supernatant solution was loaded on a High Performance Ni$^{2+}$-Sepharose column (HisTrap$^{\text{TM}}$ HP, 5 ml) equilibrated with Buffer A. Protein was eluted with Buffer A supplemented with 0.5 M imidazole. The active fractions were pooled, dialyzed against Buffer B (25 mM PIPES, pH 6.5, 5 mM MgCl$_2$, and 0.2 M KCl) and loaded on a Sephacryl$^{\text{TM}}$ S-200 high resolution column (Amersham Biosciences Co). The active fractions were pooled, concentrated and used as the purified enzyme preparation. Selenomethionine derivative PhPFK was expressed in DL41 (DE3) cells (Novagen) in minimal medium supplemented with selenomethionine and purified under the same conditions as the native protein.
A.1.6.2 Determination of Enzyme Activity

Briefly, 2-5 µL of enzyme preparation was mixed with reaction buffer containing 25 mM PIPES buffer, pH 6.5, 7 mM MgCl₂, 2 mM fructose 6-phosphate, 0.2 mM NADH, 2 mM ADP, 1.96 U α-glycerophosphate dehydrogenase, 19.6 U triosephosphate isomerase, and 0.52 U aldolase in a final volume of 0.7 mL at 50°C (standard assay). The absorbance of NADH was measured at 340 nm (ε = 6.22 mM⁻¹ cm⁻¹). Specific enzyme activities were calculated from initial linear rates and expressed in units mg⁻¹ of protein. One unit was defined as the amount of enzyme required to convert 1 µmol of F6P min⁻¹.

A.1.6.3 Hydrodynamic Radius

The hydrodynamic radius of PhPFK was determined by dynamic light scattering using a DLS DynaPro MSTC014 system (Protein Solutions Inc.) 0.52 mg/mL of enzyme was incubated in 7 mM MgCl₂, 8% glycerol and 25 mM PIPES, pH 6.5 in a final volume of 0.03 mL in a quartz-cuvette. Measurements were done at 40, 45, 50, 55, and 60°C after 20 min of incubation, and recorded at least 30 times in triplicate for each temperature. The data was processed using the DYNAMICS program (Protein Solutions Inc.)

All ADP-PFKs have a molecular weight between 50 and 54 kDa. When purified from endogenous sources, ADP-PFK from Archaeoglobus fulgidus has been reported to be dimer in solution with the presence of some tetramer at higher concentrations (140). The ADP-PFK from T. litoralis has also been reported as a homodimer (141). However, when heterologously expressed in E. coli, the ADP-PFKs from Thermococcus zilligii and
*P. furiosus* behave as tetramers. The latter of which is a special case, since when it was purified from its endogenous source it was not possible to determine its molecular weight, but it was definitely not a tetramer (111;142;143). The molecular weight of *PhPFK* expressed in *E. coli* was determined at 40, 45, 50, 55, and 60°C using dynamic light scattering and summarized in Table A.1-3. The hydrodynamic radius of *PhPFK* varies directly with temperature ranging from $5.67 \pm 0.31$ nm at 40°C to $6.22 \pm 0.21$ nm at 60°C, which corresponds to a tetramer of *PhPFK*, based on the calculated monomer molecular weight of approximately 52 kDa. However, it was not possible to generate a tetramer by applying any crystallographic symmetry transformation on the dimer of *PhPFK* found in the asymmetric unit. Moreover, when the crystal structure is analyzed with the Protein Quaternary Structure server from EMBL the protein appears to be a monomer. The dimer interface seems to be formed by nonspecific contacts produced by the crystal packing. Notably, this enzyme has a great tendency to form aggregates in solution suggesting that the hydrodynamic radius seen by DLS could be strongly influenced by this phenomenon even when the large aggregates were removed from data prior to the analysis. Also, as mentioned before, the heterologous expression system seems to affect the aggregation state of some enzymes in the family (as seen in *T. zilligii* ADP-PFK). However, considering that there is neither a need for a specific aggregation state in the family as judged from the oligomeric state of other ADP-PFK nor a sign for allosterism or any intersubunit communication (suggesting the need for a high aggregation state) and taking into account the structural information given before, it is possible that *PhPFK* may be a monomer in vivo. Yet, this issue waits for a direct
demonstration by purifying the enzyme from an endogenous source, but that goes beyond the scope of this article.

A.1.6.4 Molecular Modeling

To model the interaction of the protein with the sugar the primary sequence of *Ph*PFK was threaded onto the *Pf*GK structure bound to both AMP and glucose using the Phyre server (95), and the resulting model was energy minimized. Then F6P was docked with the structure using AutoDock (136). Briefly, the threading model with the docked F6P was simulated with the NAMD 2.6 program (135). The system was solvated with a water box at least 12 Å away from the last protein atom in each direction. The system was neutralized with NaCl ions to a final concentration of 0.1 M. The simulation was performed using a 2 fs timestep where the hydrogen position were restrained using the SHAKE algorithm. Temperature was held constant at 320 K by Langevin dynamics using a damping coefficient of 5 ps⁻¹. Pressure was maintained at 1 atm using a Nése-Hoover Langevin piston. Simulation was equilibrated for 1 ns and data was then collected for 1 ns. The CHARMM force field was used for protein, water and ions (136), while an approximation for the parameters of F6P was used as described previously (137).

The bifunctional A71E mutant *Ph*PFK bound to glucose and AMP was also modeled by threading the primary sequence on the AMP and glucose bound *Pf*GK structure using the Phyre server (95). However, instead of docking glucose, the glucose position from the *Pf*GK structure was used.
Table A.1-3 Dynamic light scattering measurements of *Ph*PFK.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Hydrodynamic radius</th>
<th>Molecular Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>°C</td>
<td>nm</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>5.67 ± 0.31</td>
<td>196,000 ± 26,000</td>
</tr>
<tr>
<td>45</td>
<td>5.74 ± 0.30</td>
<td>202,000 ± 25,000</td>
</tr>
<tr>
<td>50</td>
<td>5.73 ± 0.25</td>
<td>200,000 ± 21,000</td>
</tr>
<tr>
<td>55</td>
<td>5.88 ± 0.32</td>
<td>213,000 ± 28,000</td>
</tr>
<tr>
<td>60</td>
<td>6.22 ± 0.21</td>
<td>243,000 ± 19,000</td>
</tr>
</tbody>
</table>
Figure A.1-7 Molecular dynamics study of \textit{PhPFK-F6P} docked conformation. Average of several frames of the molecular dynamics simulation performed with the \textit{PhPFK-F6P} using the docked conformation as a starting structure.
Figure A.1-8 Model of AMP and glucose bound bifunctional (A71E) PhPFK. The mutated residue, highlighted using red text, forms a critical hydrogen bond with the C2 hydroxyl group of glucose, which allows the enzyme to both bind and phosphorylate glucose in addition to F6P.